(11) **EP 0 939 818 B1** 

#### (12)

#### **EUROPEAN PATENT SPECIFICATION**

- (45) Date of publication and mention of the grant of the patent: 27.04.2005 Bulletin 2005/17
- (21) Application number: 97942094.0
- (22) Date of filing: 22.08.1997

- (51) Int Cl.<sup>7</sup>: **C12N 15/31**, C12N 1/21, C12P 21/02, C07K 14/33, A61K 38/16, A61K 39/08
- (86) International application number: PCT/GB1997/002273
- (87) International publication number: WO 1998/007864 (26.02.1998 Gazette 1998/08)

### (54) RECOMBINANT TOXIN FRAGMENTS

REKOMBINANTE TOXINFRAGMENTE FRAGMENTS DE TOXINES RECOMBINES

- (84) Designated Contracting States:

  AT BE CH DE DK FR GB IT LI NL SE
- (30) Priority: **23.08.1996 GB 9617671 13.12.1996 GB 9625996**
- (43) Date of publication of application: 08.09.1999 Bulletin 1999/36
- (73) Proprietors:
  - Health Protection Agency Salisbury, Wiltshire SP4 0JG (GB)
  - Ipsen Limited
     Slough Berkshire SL1 3XE (GB)
- (72) Inventors:
  - SHONE, Clifford, Charles, CAMR
     Porton Down, Salisbury, Wiltshire SP4 0JG (GB)
  - QUINN, Conrad, Padraig, CAMR
     Porton Down, Salisbury, Wiltshire SP4 0JG (GB)
  - FOSTER, Keith, Alan, CAMR
     Porton Down, Salisbury, Wiltshire SP4 0JG (GB)
- (74) Representative: MacLean, Martin Robert Mathys & Squire 120 Holborn London EC1N 2SQ (GB)

- (56) References cited: WO-A-96/12802
  - KURAZONO H ET AL: "Minimal essential \*domains\* specifying toxicity of the \*light\* \*chains\* of tetanus toxin and botulinum neurotoxin type A." J BIOL CHEM, JUL 25 1992, 267 (21) P14721-9, UNITED STATES, XP002047910
  - LI ET AL: "A SINGLE MUTATION IN THE RECOMBINANT LIGHT CHAIN OF TETANUS TOXIN ABOLISHES ITS PROTEOLYTIC ACTIVITY AND REMOVES THE TOXICITY SEEN AFTER RECONSTITUTION WITH NATIVE HEAVY CHAIN" BIOCHEMISTRY, vol. 33, no. 22, 1994, pages 7014-7020, XP002015938
  - BINZ T ET AL: "THE COMPLETE SEQUENCE OF BOTULINUM NEUROTOXIN TYPE A AND COMPARISON WITH OTHER CLOSTRIDIAL NEUROTOXINS" JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 265, no. 16, 5 June 1990, pages 9153-9158, XP002009348

### Remarks:

The file contains technical information submitted after the application was filed and not included in this specification

P 0 939 818 B

Note: Within nine months from the publication of the mention of the grant of the European patent, any person may give notice to the European Patent Office of opposition to the European patent granted. Notice of opposition shall be filed in a written reasoned statement. It shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

#### Description

10

15

25

30

35

40

45

[0001] This invention relates to recombinant toxin fragments, to DNA encoding these fragments and to their uses such as in a vaccine and for *in vitro* and *in vivo* purposes.

[0002] The clostridial neurotoxins are potent inhibitors of calcium-dependent neurotransmitter secretion in neuronal cells. They are currently considered to mediate this activity through a specific endoproteolytic cleavage of at least one of three vesicle or pre-synaptic membrane associated proteins VAMP, syntaxin or SNAP-25 which are central to the vesicle docking and membrane fusion events of neurotransmitter secretion. The neuronal cell targeting of tetanus and botulinum neurotoxins is considered to be a receptor mediated event following which the toxins become internalised and subsequently traffic to the appropriate intracellular compartment where they effect their endopeptidase activity.

[0003] Inhibition of calcium-independent [3H] noradrenaline outflow from freeze-thawed synaptosomes has also been reported (Hausinger, A. et al 1995, Toxicon, vol. 33, No. 11, p. 1519 - 1530)

[0004] The clostridial neurotoxins share a common architecture of a catalytic L-chain (LC, ca 50 kDa) disulphide linked to a receptor binding and translocating H-chain (HC, ca 100 kDa). The HC polypeptide is considered to comprise all or part of two distinct functional domains. The carboxy-terminal half of the HC (ca 50 kDa), termed the  $H_C$  domain, is involved in the high affinity, neurospecific binding of the neurotoxin to cell surface receptors on the target neuron, whilst the amino-terminal half, termed the  $H_N$  domain (ca 50 kDa), is considered to mediate the translocation of at least some portion of the neurotoxin across cellular membranes such that the functional activity of the LC is expressed within the target cell. The  $H_N$  domain also, has the property, under conditions of low pH, of forming ion-permeable channels in lipid membranes, this may in some manner relate to its translocation function.

[0005] For botulinum neurotoxin type A (BoNT/A) these domains are considered to reside within amino acid residues 872-1296 for the  $H_C$ , amino acid residues 449-871 for the  $H_N$  and residues 1-448 for the LC. Digestion with trypsin effectively degrades the  $H_C$  domain of the BoNT/A to generate a non-toxic fragment designated  $LH_N$ , which is no longer able to bind to and enter neurons (Fig. 1). The  $LH_N$  fragment so produced also has the property of enhanced solubility compared to both the parent holotoxin and the isolated LC.

[0006] It is therefore possible to provide functional definitions of the domains within the neurotoxin molecule, as follows:

(A) clostridial neurotoxin light chain:

- a metalloprotease exhibiting high substrate specificity for vesicle and/or plasma membrane associated proteins involved in the exocytotic process. In particular, it cleaves one or more of SNAP-25, VAMP (synaptobrevin / cellubrevin) and syntaxin. A single mutation (Glu<sup>224</sup> → Ala<sup>234</sup>) in the light chain of tetanus toxin has been shown to abolish proteolytic activity (Li, Y. et al, Biochemistry 1994, 33, p. 7014 7020).
- (B) clostridial neurotoxin heavy chain H<sub>N</sub> domain:
- a portion of the heavy chain which enables translocation of that portion of the neurotoxin molecule such that a functional expression of light chain activity occurs within a target cell.
- the domain responsible for translocation of the endopeptidase activity, following binding of neurotoxin to its specific cell surface receptor via the binding domain, into the target cell.
- the domain responsible for formation of ion-permeable pores in lipid membranes under conditions of low pH.
- the domain responsible for increasing the solubility of the entire polypeptide compared to the solubility of light chain alone.
- (C) clostridial neurotoxin heavy chain H<sub>C</sub> domain.
- a portion of the heavy chain which is responsible for binding of the native holotoxin to cell surface receptor(s) involved in the intoxicating action of clostridial toxin prior to internalisation of the toxin into the cell.

[0007] The identity of the cellular recognition markers for these toxins is currently not understood and no specific receptor species have yet been identified although Kozaki et al. have reported that synaptotagmin may be the receptor for botulinum neurotoxin type B. It is probable that each of the neurotoxins has a different receptor.

[0008] It is desirable to have positive controls for toxin assays, to develop clostridial toxin vaccines and to develop therapeutic agents incorporating desirable properties of clostridial toxin.

[0009] For example, WO96/12802 describes vaccine compositions comprising the H<sub>C</sub> region of C. botulinum.

[0010] However, due to its extreme toxicity, the handling of native toxin is hazardous.

[0011] The present invention seeks to overcome or at least ameliorate problems associated with production and handling of clostridial toxin.

[0012] Accordingly, the invention provides a polypeptide as defined in Claim 1.

20

25

30

**[0013]** The invention may thus provide a single polypeptide chain containing a domain equivalent to a clostridial toxin light chain and a domain providing the functional aspects of the  $H_N$  of a clostridial toxin heavy chain, whilst lacking the functional aspects of a clostridial toxin  $H_C$  domain.

[0014] For the purposes of the invention, the functional property or properties of the  $H_N$  of a clostridial toxin heavy chain that are required to be exhibited by the second domain of the polypeptide of the invention are either (i) translocation of the polypeptide into a cell, or (ii) increasing solubility of the polypeptide compared to solubility of the first domain on its own or (iii) both (i) and (ii). References hereafter to a  $H_N$  domain or to the functions of a  $H_N$  domain are references to this property or properties. The second domain is not required to exhibit other properties of the  $H_N$  domain of a clostridial toxin heavy chain.

[0015] A polypeptide of the invention can thus be soluble but lack the translocation function of a native toxin-this is of use in providing an immunogen for vaccinating or assisting to vaccinate an individual against challenge by toxin. In a specific embodiment of the invention described in an example below a polypeptide designated LH<sub>423</sub>/A elicited neutralising antibodies against type A neurotoxin. A polypeptide of the invention can likewise thus be relatively insoluble but retain the translocation function of a native toxin - this is of use if solubility is imparted to a composition made up of that polypeptide and one or more other components by one or more of said other components.

[0016] The first domain of the polypeptide of the invention cleaves one or more vesicle or plasma-membrane associated proteins essential to the specific cellular process of exocytosis, and cleavage of these proteins results in inhibition of exocytosis, typically in a non-cytotoxic manner. The cell or cells affected are not restricted to a particular type or subgroup but can include both neuronal and non-neuronal cells. The activity of clostridial neurotoxins in inhibiting exocytosis has indeed, been observed almost universally in eukaryotic cells expressing a relevant cell surface receptor, including such diverse cells as from Aplysia (sea slug), Drosophila (fruit fly) and mammalian nerve cells, and the activity of the first domain is to be understood as including a corresponding range of cells.

[0017] The polypeptide of the invention may be obtained by expression of a recombinant nucleic acid, preferably a DNA, and is a single polypeptide, that is to say not cleaved into separate light and heavy chain domains. The polypeptide is thus available in convenient and large quantities using recombinant techniques.

[0018] In a polypeptide according to the invention, said first domain preferably comprises a clostridial toxin light chain or a fragment or variant of a clostridial toxin light chain. The fragment is optionally an N-terminal, or C-terminal fragment of the light chain, or is an internal fragment, so long as it substantially retains the ability to cleave the vesicle or plasmamembrane associated protein essential to exocytosis. The minimal domains necessary for the activity of the light chain of clostridial toxins are described in J. Biol. Chem., Vol.267, No. 21, July 1992, pages 14721-14729. The variant has a different peptide sequence from the light chain or from the fragment, though it too is capable of cleaving the vesicle or plasma-membrane associated protein. It is conveniently obtained by insertion, deletion and/or substitution of a light chain or fragment thereof. In embodiments of the invention described below a variant sequence comprises (i) an N-terminal extension to a clostridial toxin light chain or fragment modified by alteration of at least one amino acid (iii) a C-terminal extension to a clostridial toxin light chain or fragment, or (iv) combinations of 2 or more of (i)-(iii).

[0019] In further embodiments of the invention, the variant contains an amino acid sequence modified so that (a) there is no protease sensitive region between the LC and  $H_N$  components of the polypeptide, or (b) the protease sensitive region is specific for a particular protease. This latter embodiment is of use if it is desired to activate the endopeptidase activity of the light chain in a particular environment or cell. Though, in general, the polypeptides of the invention are activated prior to administration.

[0020] The first domain preferably exhibits endopeptidase activity specific for a substrate selected from one or more of SNAP-25, synaptobrevin/VAMP and syntaxin. The clostridial toxin is preferably botulinum toxin or tetanus toxin.

[0021] In an embodiment of the invention described in an example below, the toxin light chain and the portion of the toxin heavy chain are of botulinum toxin type A. In a further embodiment of the invention described in an example below, the toxin light chain and the portion of the toxin heavy chain are of botulinum toxin type B. The polypeptide optionally comprises a light chain or fragment or variant of one toxin type and a heavy chain or fragment or variant of another toxin type.

[0022] In a polypeptide according to the invention said second domain preferably comprises a clostridial toxin heavy chain  $H_N$  portion or a fragment or variant of a clostridial toxin heavy chain  $H_N$  portion. The fragment is optionally an N-terminal or C-terminal or internal fragment, so long as it retains the function of the  $H_N$  domain. Teachings of regions within the  $H_N$  responsible for its function are provided for example in Biochemistry 1995, 34, pages 15175-15181 and Eur. J. Biochem, 1989, 185, pages 197-203. The variant has a different sequence from the  $H_N$  domain or fragment,

though it too retains the function of the  $H_N$  domain. It is conveniently obtained by insertion, deletion and/or substitution of a  $H_N$  domain or fragment thereof. In embodiments of the invention, described below, it comprises (i) an N-terminal extension to a  $H_N$  domain or fragment, (ii) a C-terminal extension to a  $H_N$  domain or fragment, (iii) a modification to a  $H_N$  domain or fragment by alteration of at least one amino acid, or (iv) combinations of 2 or more of (i)-(iii). The clostridial toxin is preferably botulinum toxin or tetanus toxin.

[0023] The invention also provides a polypeptide comprising a clostridial neurotoxin light chain and a N-terminal fragment of a clostridial neurotoxin heavy chain, the fragment preferably comprising at least 423 of the N-terminal amino acids of the heavy chain of botulinum toxin type A, 417 of the N-terminal amino acids of the heavy chain of botulinum toxin type B or the equivalent number of N-terminal amino acids of the heavy chain of other types of clostridial toxin such that the fragment possesses an equivalent alignment of homologous amino acid residues.

**[0024]** These polypeptides of the invention are thus not composed of two or more polypeptides, linked for example by di-sulphide bridges into composite molecules. Instead, these polypeptides are single chains and are not active or their activity is significantly reduced in an *in vitro* assay of neurotoxin endopeptidase activity.

[0025] Further, the polypeptides may be susceptible to be converted into a form exhibiting endopeptidase activity by the action of a proteolytic agent, such as trypsin. In this way it is possible to control the endopeptidase activity of the toxin light chain.

[0026] In a specific embodiment of the invention described in an example below, there is provided a polypeptide lacking a portion designated H<sub>C</sub> of a clostridial toxin heavy chain. This portion, seen in the naturally produced toxin, is responsible for binding of toxin to cell surface receptors prior to internalisation of the toxin. This specific embodiment is therefore adapted so that it can not be converted into active toxin, for example by the action of a proteolytic enzyme. The invention thus also provides a polypeptide comprising a clostridial toxin light chain and a fragment of a clostridial toxin heavy chain, said fragment being not capable of binding to those cell surface receptors involved in the intoxicating action of clostridial toxin, and it is preferred that such a polypeptide lacks an intact portion designated H<sub>C</sub> of a clostridial toxin heavy chain.

[0027] In further embodiments of the invention there are provided compositions containing a polypeptide comprising a clostridial toxin light chain and a portion designated H<sub>N</sub> of a clostridial toxin heavy chain, and wherein the composition is free of clostridial toxin and free of any clostridial toxin precursor that may be converted into clostridial toxin by the action of a proteolytic enzyme. Examples of these compositions include those containing toxin light chain and H<sub>N</sub> sequences of botulinum toxin types A, B, C, D, E, F and G.

30

[0028] The polypeptides of the invention are conveniently adapted to bind to, or include, a ligand for targeting to desired cells. The polypeptide optionally comprises a sequence that binds to, for example, an immunoglobulin. A suitable sequence is a tandem repeat synthetic IgG binding domain derived from domain B of Staphylococcal protein A. Choice of immunoglobulin specificity then determines the target for a polypeptide - immunoglobulin complex. Alternatively, the polypeptide comprises a non-clostridial sequence that binds to a cell surface receptor, suitable sequences including insulin-like growth factor-1 (IGF-1) which binds to its specific receptor on particular cell types and the 14 amino acid residue sequence from the carboxy-terminus of cholera toxin A subunit which is able to bind the cholera toxin B subunit and thence to GM1 gangliosides. A polypeptide according to the invention thus, optionally, further comprises a third domain adapted for binding of the polypeptide to a cell.

[0029] In a second aspect the invention provides a fusion protein comprising a fusion of (a) a polypeptide of the invention as described above with (b) a second polypeptide adapted for binding to a chromatography matrix so as to enable purification of the fusion protein using said chromatography matrix. It is convenient for the second polypeptide to be adapted to bind to an affinity matrix, such as a glutathione Sepharose, enabling rapid separation and purification of the fusion protein from an impure source, such as a cell extract or supernatant.

[0030] One possible second purification polypeptide is glutathione-S-transferase (GST), and others will be apparent to a person of skill in the art, being chosen so as to enable purification on a chromatography column according to conventional techniques.

[0031] As noted above, by proteolytic treatment, for example using trypsin, of a polypeptide of the invention it is possible to induce endopeptidase activity in the treated polypeptide.

[0032] A third aspect of the invention provides a composition comprising a polypeptide of the present invention, said composition being non-toxic *in vivo*. The overall endopeptidase activity of the composition will, of course, be determined by the amount of the polypeptide that is present.

[0033] While it is known to treat naturally produced clostridial toxin to remove the  $H_C$  domain, this treatment does not totally remove toxicity of the preparation, instead some residual toxin activity remains. Natural toxin treated in this way is therefore still not entirely safe. The composition of the invention, derived by treatment of a pure source of polypeptide advantageously is free of toxicity, and can conveniently be used as a positive control in a toxin assay, as a vaccine against clostridial toxin or for other purposes where it is essential that there is no residual toxicity in the composition.

[0034] The invention enables production of the polypeptides and fusion proteins of the invention by recombinant

means.

20

[0035] A fourth aspect of the invention provides a nucleic acid encoding a polypeptide or a fusion protein according to any of the aspects of the invention described above.

[0036] In one embodiment of this aspect of the invention, a DNA sequence provided to code for the polypeptide or fusion protein is not derived from native clostridial sequences, but is an artificially derived sequence not preexisting in nature.

[0037] A specific DNA (SEQ ID NO: 1) described in more detail below encodes a polypeptide or a fusion protein comprising nucleotides encoding residues 1-871 of a botulinum toxin type A. Said polypeptide comprises the light chain domain and the first 423 amino acid residues of the amino terminal portion of a botulinum toxin type A heavy chain. This recombinant product is designated LH $_{423}$ /A (SEQ ID NO: 2).

[0038] In a second embodiment of this aspect of the invention a DNA sequence which codes for the polypeptide or fusion protein is derived from native clostridial sequences but codes for a polypeptide or fusion protein not found in nature.

[0039] A specific DNA (SEQ ID NO: 19) described in more detail below encodes a polypeptide or a fusion protein and comprises nucleotides encoding residues 1-1171 of a botulinum toxin type B. Said polypeptide comprises the light chain domain and the first 728 amino acid residues of the amino terminal protein of a botulinum type B heavy chain. This recombinant product is designated LH<sub>728</sub>/B (SEQ ID NO: 20).

**[0040]** The invention thus also provides a method of manufacture of a polypeptide comprising expressing in a host cell a DNA according to the third aspect of the invention. The host cell is suitably not able to cleave a polypeptide or fusion protein of the invention so as to separate light and heavy toxin chains; for example, a non-clostridial host.

[0041] The invention further provides a method of manufacture of a polypeptide comprising expressing in a host cell a DNA encoding a fusion protein as described above, purifying the fusion protein by elution through a chromatography column adapted to retain the fusion protein, eluting through said chromatography column a ligand adapted to displace the fusion protein and recovering the fusion protein. Production of substantially pure fusion protein is thus made possible. Likewise, the fusion protein is readily cleaved to yield a polypeptide of the invention, again in substantially pure form, as the second polypeptide may conveniently be removed using the same type of chromatography column.

[0042] The  $LH_N/A$  derived from dichain native toxin requires extended digestion with trypsin to remove the C-terminal half of the heavy chain, the  $H_C$  domain. The loss of this domain effectively renders the toxin inactive *in vivo* by preventing its interaction with host target cells. There is, however, a residual toxic activity which may indicate a contaminating, trypsin insensitive, form of the whole type A neurotoxin.

[0043] In contrast, the recombinant preparations of the invention are the product of a discreet, defined gene coding sequence and can not be contaminated by full length toxin protein. Furthermore, the product as recovered from *E.coli*, and from other recombinant expression hosts, is an inactive single chain peptide, or if expression hosts produce a processed, active polypeptide it is not a toxin. Endopeptidase activity of LH<sub>423</sub>/A, as assessed by the current *in vitro* peptide cleavage assay, is wholly dependent on activation of the recombinant molecule between residues 430 and 454 by trypsin. Other proteolytic enzymes that cleave between these two residues are generally also suitable for activation of the recombinant molecule. Trypsin cleaves the peptide bond C-terminal to Arginine or C-terminal to Lysine and is suitable as these residues are found in the 430-454 region and are exposed (see Fig. 12).

[0044] The recombinant polypeptides of the invention are potential therapeutic agents for targeting to cells expressing the relevant substrate but which are not implicated in effecting botulism. An example might be where secretion of neurotransmitter is inappropriate or undesirable or alternatively where a neuronal cell is hyperactive in terms of regulated secretion of substances other than neurotransmitter. In such an example the function of the H<sub>C</sub> domain of the native toxin could be replaced by an alternative targeting sequence providing, for example, a cell receptor ligand and/or translocation domain.

[0045] One application of the recombinant polypeptides of the invention will be as a reagent component for synthesis of therapeutic molecules, such as disclosed in WO-A-94/21300. The recombinant product will also find application as a non-toxic standard for the assessment and development of *in vitro* assays for detection of functional botulinum or tetanus neurotoxins either in foodstuffs or in environmental samples, for example as disclosed in EP-A-0763131.

[0046] A further option is addition, to the C-terminal end of a polypeptide of the invention, of a peptide sequence which allows specific chemical conjugation to targeting ligands of both protein and non-protein origin.

[0047] In yet a further embodiment an alternative targeting ligand is added to the N-terminus of polypeptides of the invention. Recombinant LH<sub>N</sub> derivatives have been designated that have specific protease cleavage sites engineered at the C-terminus of the LC at the putative trypsin sensitive region and also at the extreme C-terminus of the complete protein product. These sites will enhance the activational specificity of the recombinant product such that the dichain species can only be activated by proteolytic cleavage of a more predictable nature than use of trypsin.

[0048] The  $LH_N$  enzymatically produced from native BoNT/A is an efficient immunogen and thus the recombinant form with its total divorce from any full length neurotoxin represents a vaccine component. The recombinant product may serve as a basal reagent for creating defined protein modifications in support of any of the above areas.

[0049] Recombinant constructs are assigned distinguishing names on the basis of their amino acid sequence length and their Light Chain (L-chain, L) and Heavy Chain (H-chain, H) content as these relate to translated DNA sequences in the public domain or specifically to SEQ ID NO: 2 and SEQ ID NO: 20. The 'LH' designation is followed by '/X' where 'X' denotes the corresponding clostridial toxin serotype or class, e.g. 'A' for botulinum neurotoxin type A or 'TeTx' for tetanus toxin. Sequence variants from that of the native toxin polypeptide are given in parenthesis in standard format, namely the residue position number prefixed by the residue of the native sequence and suffixed by the residue of the variant.

[0050] Subscript number prefixes indicate an amino-terminal (N-terminal) extension, or where negative a deletion, to the translated sequence. Similarly, subscript number suffixes indicate a carboxy terminal (C-terminal) extension or where negative numbers are used, a deletion. Specific sequence inserts such as protease cleavage sites are indicated using abbreviations, e.g. Factor Xa is abbreviated to FXa. L-chain C-terminal suffixes and H-chain N-terminal prefixes are separated by a '/' to indicate the predicted junction between the L and H-chains. Abbreviations for engineered ligand sequences are prefixed or suffixed to the clostridial L-chain or H-chain corresponding to their position in the translation product.

[0051] Following this nomenclature,

15

35

40

45

50

55

	LH <sub>423</sub> /A =	SEQ ID NO: 2, containing the entire L-chain and 423 amino acids of the H-chain of botulinum neurotoxin type A;
20	<sub>2</sub> LH <sub>423</sub> /A =	a variant of this molecule, containing a two amino acid extension to the N-terminus of the L-chain;
25	$_{2}L_{/2}H_{423}/A =$	a further variant in which the molecule contains a two amino acid extension on the N-terminus of both the L-chain and the H-chain;
30	<sub>2</sub> L <sub>FXa/2</sub> H <sub>423</sub> /A =	a further variant containing a two amino acid extension to the N-terminus of the L-chain, and a Factor Xa cleavage sequence at the C-terminus of the L-chain which, after cleavage of the molecule with Factor Xa leaves a two amino acid N-terminal extension to the H-chain component; and
30	$_{2}L_{FXa/2}H_{423}/A$ -IGF-1 =	a variant of this molecule which has a further C-terminal extension to the H-chain, in this example the insulin-like growth factor 1 (IGF-1) sequence.

[0052] There now follows description of specific embodiments of the invention, illustrated by drawings in which:

Fig. 1 shows a schematic representation of the domain structure of botulinum neurotoxin type A (BoNT/A);

Fig. 2 shows a schematic representation of assembly of the gene for an embodiment of the invention designated LH<sub>423</sub>/A;

Fig.3 is a graph comparing activity of native toxin, trypsin generated "native"  $LH_N/A$  and an embodiment of the invention designated  $_2LH_{423}/A(Q_2E,N_{26}K,A_{27}Y)$  in an *in vitro* peptide cleavage assay;

Fig.4 is a comparison of the first 33 amino acids in published sequences of native toxin and embodiments of the invention;

Fig.5 shows the transition region of an embodiment of the invention designated  $L/_4H_{423}/A$  illustrating insertion of four amino acids at the N-terminus of the  $H_N$  sequence; amino acids coded for by the *Eco* 47 III restriction endonuclease cleavage site are marked and the  $H_N$  sequence then begins ALN...;

Fig.6 shows the transition region of an embodiment of the invention designated  $L_{FXa/3}H_{423}/A$  illustrating insertion of a Factor Xa cleavage site at the C-terminus of the L-chain, and three additional amino acids coded for at the N-terminus of the H-sequence; the N-terminal amino acid of the cleavage-activated  $H_N$  will be cysteine;

Fig.7 shows the C-terminal portion of the amino acid sequence of an embodiment of the invention designated  $L_{FXa/3}H_{423}/A$ -IGF-1, a fusion protein; the IGF-1 sequence begins at position  $G_{882}$ ;

Fig.8 shows the C-terminal portion of the amino acid sequence of an embodiment of the invention designated

L<sub>FXa/3</sub>H<sub>423</sub>/A-CtxA14, a fusion protein; the C-terminal CtxA sequence begins at position Q<sub>882</sub>;

Fig.9 shows the C-terminal portion of the amino acid sequence of an embodiment of the invention designated  $L_{FXa/3}H_{423}/A$ -ZZ, a fusion protein; the C-terminal ZZ sequence begins at position  $A_{890}$  immediately after a genenase recognition site (underlined);

Figs.10 & 11 show schematic representations of manipulations of polypeptides of the invention; Fig.10 shows LH<sub>423</sub>/A with N-terminal addition of an affinity purification peptide (in this case GST) and C-terminal addition of an Ig binding domain; protease cleavage sites R1, R2 and R3 enable selective enzymatic separation of domains; Fig. 11 shows specific examples of protease cleavage sites R1, R2 and R3 and a C-terminal fusion peptide sequence;

Fig.12 shows the trypsin sensitive activation region of a polypeptide of the invention;

5

10

15

20

Fig.13 shows Western blot analysis of recombinant LH<sub>107</sub>/B expressed from *E.coli*; panel A was probed with anti-BoNT/B antiserum; Lane 1, molecular weight standards; lanes 2 & 3, native BoNT/B; lane 4, immunopurified LH<sub>107</sub>/B; panel B was probed with anti-T7 peptide tag antiserum; lane 1, molecular weight standards; lanes 2 & 3, positive control *E.coli* T7 expression; lane 4 immunopurified LH<sub>107</sub>/B.

[0053] The sequence listing that accompanies this application contains the following sequences:-

	SEQ ID NO:	SEQUENCE
	1	DNA coding for LH <sub>423</sub> /A
	2	LH <sub>423</sub> /A
25		
	3	DNA coding for <sub>23</sub> LH <sub>423</sub> /A (Q <sub>2</sub> E,N <sub>26</sub> K,A <sub>27</sub> Y), of which an N-terminal portion is shown in Fig. 4.
	4	<sub>23</sub> LH <sub>423</sub> /A (Q <sub>2</sub> E,N <sub>26</sub> K,A <sub>27</sub> Y)
	5	DNA coding for <sub>2</sub> LH <sub>423</sub> /A (Q <sub>2</sub> E,N <sub>26</sub> K,A <sub>27</sub> Y), of which an N-terminal portion is shown in Fig.4
20	6	<sub>2</sub> LH <sub>423</sub> /A (Q <sub>2</sub> E,N <sub>26</sub> K,A <sub>27</sub> Y)
30		
	7	DNA coding for native BoNT/A according to Binz et al
	8	native BoNT/A according to Binz et al
	9	DNA coding for L <sub>/4</sub> H <sub>423</sub> /A
35	10	L <sub>/4</sub> H <sub>423</sub> /A
	11	DNA coding for L <sub>FXa</sub> / <sub>3</sub> H <sub>423</sub> /A
	12	L <sub>FXa</sub> / <sub>3</sub> H <sub>423</sub> /A
	13	DNA coding for L <sub>FXa</sub> / <sub>3</sub> H <sub>423</sub> /A-IGF-1
40	14	L <sub>FXa</sub> / <sub>3</sub> H <sub>423</sub> /A-IGF-1
40	15	DNA coding for L <sub>FXe</sub> / <sub>3</sub> H <sub>423</sub> /A-CtxA14
	16	L <sub>FXa</sub> / <sub>3</sub> H <sub>423</sub> /A-CtxA 14
1	17	DNA coding for L <sub>FXa/3</sub> H <sub>423</sub> /A-ZZ
	18	L <sub>FXa/3</sub> H <sub>423</sub> /A-ZZ
45		
	19	DNA coding for LH <sub>728</sub> /B
	20	LH <sub>728</sub> /B
	21	DNA coding for LH <sub>417</sub> /B
-	22	LH <sub>417</sub> /B
50	23	DNA coding for LH <sub>107</sub> /B
	24	LH <sub>107</sub> /B
	25	DNA coding for LH <sub>423</sub> /A (Q <sub>2</sub> E,N <sub>26</sub> K,A <sub>27</sub> Y)
	26	$LH_{423}/A (Q_2E, N_{26}K, A_{27}Y)$
55	27	DNA coding for LH <sub>417</sub> /B wherein the first 274 bases are modified to have an <i>E.coli</i> codon bias
	28	DNA coding for LH <sub>417</sub> /B wherein bases 691-1641 of the native BoNT/B sequence have been
		replaced by a degenerate DNA coding for amino acid residues 231-547 of the native BoNT/B polypeptide
L	L	polypapilide

#### Example 1

10

20

25

45

[0054] A 2616 base pair, double stranded gene sequence (SEQ ID NO: 1) has been assembled from a combination of synthetic, chromosomal and polymerase-chain-reaction generated DNA (Figure 2). The gene codes for a polypeptide of 871 amino acid residues corresponding to the entire light-chain (LC, 448 amino acids) and 423 residues of the amino terminus of the heavy-chain ( $H_C$ ) of botulinum neurotoxin type A. This recombinant product is designated the LH<sub>423</sub>/A fragment (SEQ ID NO: 2).

#### Construction of the recombinant product

[0055] The first 918 base pairs of the recombinant gene were synthesised by concatenation of short oligonucleotides to generate a coding sequence with an E. coli codon bias. Both DNA strands in this region were completely synthesised as short overlapping oligonucleotides which were phosphorylated, annealed and ligated to generate the full synthetic region ending with a unique Kpnl restriction site. The remainder of the  $LH_{423}/A$  coding sequence was PCR amplified from total chromosomal DNA from Clostridium botulinum and annealed to the synthetic portion of the gene.

[0056] The internal PCR amplified product sequences were then deleted and replaced with the native, fully sequenced, regions from clones of *C. botulinum* chromosomal origin to generate the final gene construct. The final composition is synthetic DNA (bases 1-913), polymerase amplified DNA (bases 914-1138 and 1976-2616) and the remainder is of *C. botulinum* chromosomal origin (bases 1139-1975). The assembled gene was then fully sequenced and cloned into a variety of *E.coli* plasmid vectors for expression analysis.

### Expression of the recombinant gene and recovery of protein product

[0057] The DNA is expressed in *E. coli* as a single nucleic acid transcript producing a soluble single chain polypeptide of 99,951 Daltons predicted molecular weight. The gene is currently expressed in *E. coli* as a fusion to the commercially available coding sequence of glutathione S-transferase (GST) of *Schistosoma japonicum* but any of an extensive range of recombinant gene expression vectors such as pEZZ18, pTrc99, pFLAG or the pMAL series may be equally effective as might expression in other prokaryotic or eukaryotic hosts such as the Gram positive bacilli, the yeast *P. pastoris* or in insect or mammalian cells under appropriate conditions.

[0058] Currently, *E. coli* harbouring the expression construct is grown in Luria-Bertani broth (L-broth pH 7.0, containing 10 g/l bacto-tryptone, 5 g/l bacto-yeast extract and 10 g/l sodium chloride) at 37° C until the cell density (biomass) has an optical absorbance of 0.4- 0.6 at 600 nm and the cells are in mid-logarithmic growth phase. Expression of the gene is then induced by addition of isopropylthio-β-D-galactosidase (IPTG) to a final concentration of 0.5 mM. Recombinant gene expression is allowed to proceed for 90 min at a reduced temperature of 25°C. The cells are then harvested by centrifugation, are resuspended in a buffer solution containing 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 0.5 M NaCl, 10 mM EGTA, 0.25% Tween, pH 7.0 and then frozen at -20°C. For extraction of the recombinant protein the cells are disrupted by sonication. The cell extract is then cleared of debris by centrifugation and the cleared supernatant fluid containing soluble recombinant fusion protein (GST- LH<sub>423</sub>/A) is stored at -20°C pending purification. A proportion of recombinant material is not released by the sonication procedure and this probably reflects insolubility or inclusion body formation. Currently we do not extract this material for analysis but if desired this could be readily achieved using methods known to those skilled in the art

[0059] The recombinant GST- LH<sub>423</sub>/A is purified by adsorption onto a commercially prepared affinity matrix of glutathione Sepharose and subsequent elution with reduced glutathione. The GST affinity purification marker is then removed by proteolytic cleavage and reabsorption to glutathione Sepharose; recombinant LH<sub>423</sub>/A is recovered in the non-adsorbed material.

#### Construct variants

[0060] A variant of the molecule, LH $_{423}$ /A (Q $_2$ E,N $_{26}$ K,A $_{27}$ Y) (SEQ ID NO: 26) has been produced in which three amino acid residues have been modified within the light chain of LH $_{423}$ /A producing a polypeptide containing a light chain sequence different to that of the published amino acid sequence of the light chain of BoNT/A.

[0061] Two further variants of the gene sequence that have been expressed and the corresponding products purified are  $_{23}\text{LH}_{423}/A$  ( $Q_2\text{E},N_{26}\text{K},A_{27}\text{Y}$ ) (SEQ ID NO: 4) which has a 23 amino acid N-terminal extension as compared to the predicted native L-chain of BoNT/A and  $_2\text{LH}_{423}/A$  ( $Q_2\text{E},N_{26}\text{K},A_{27}\text{Y}$ ) (SEQ ID NO: 6) which has a 2 amino acid N-terminal extension (Figure 4).

[0062] In yet another variant a gene has been produced which contains a *Eco* 47 III restriction site between nucleotides 1344 and 1345 of the gene sequence given in (SEQ ID NO: 1). This modification provides a restriction site at the position in the gene representing the interface of the heavy and light chains in native neurotoxin, and provides the

capability to make insertions at this point using standard restriction enzyme methodologies known to those skilled in the art. It will also be obvious to those skilled in the art that any one of a number of restriction sites could be so employed, and that the *Eco* 47 III insertion simply exemplifies this approach. Similarly, it would be obvious for one skilled in the art that insertion of a restriction site in the manner described could be performed on any gene of the invention. The gene described, when expressed, codes for a polypeptide, L/4H423/A (SEQ ID NO: 10), which contains an additional four amino acids between amino acids 448 and 449 of LH423/A at a position equivalent to the amino terminus of the heavy chain of native BoNT/A.

[0063] A variant of the gene has been expressed,  $L_{FXa/3}H_{423}/A$  (SEQ ID NO: 12), in which a specific proteolytic cleavage site was incorporated at the carboxy-terminal end of the light chain domain, specifically after residue 448 of  $L_{/4}H_{423}/A$ . The cleavage site incorporated was for Factor Xa protease and was coded for by modification of SEQ ID NO: 1. It will be apparent to one skilled in the art that a cleavage site for another specified protease could be similarly incorporated, and that any gene sequence coding for the required cleavage site could be employed. Modification of the gene sequence in this manner to code for a defined protease site could be performed on any gene of the invention. [0064] Variants of  $L_{FXa/3}H_{423}/A$  have been constructed in which a third domain is present at the carboxy-terminal end of the polypeptide which incorporates a specific binding activity into the polypeptide.

[0065] Specific examples described are:

20

25

- (1) L<sub>FXa/3</sub>H<sub>423</sub>/A-IGF-1 (SEQ ID NO: 14), in which the carboxy-terminal domain has a sequence equivalent to that of insulin-like growth factor-1 (IGF-1) and is able to bind to the insulin-like growth factor receptor with high affinity;
- (2)  $L_{FXa/3}H_{423}/A$ -CtxA14 (SEQ ID NO: 16), in which the carboxy-terminal domain has a sequence equivalent to that of the 14 amino acids from the carboxy-terminus of the A-subunit of cholera toxin (CtxA) and is thereby able to interact with the cholera toxin B-subunit pentamer; and
- (3)  $L_{FXa/3}H_{423}/A$ -ZZ (SEQ ID NO: 18), in which the carboxy-terminal domain is a tandem repeating synthetic IgG binding domain. This variant also exemplifies another modification applicable to the current invention, namely the inclusion in the gene of a sequence coding for a protease cleavage site located between the end of the clostridial heavy chain sequence and the sequence coding for the binding ligand. Specifically in this example a sequence is inserted at nucleotides 2650 to 2666 coding for a genenase cleavage site. Expression of this gene produces a polypeptide which has the desired protease sensitivity at the interface between the domain providing  $H_N$  function and the binding domain. Such a modification enables selective removal of the C-terminal binding domain by treatment of the polypeptide with the relevant protease.
- [0066] It will be apparent that any one of a number of such binding domains could be incorporated into the polypeptide sequences of this invention and that the above examples are merely to exemplify the concept. Similarly, such binding domains can be incorporated into any of the polypeptide sequences that are the basis of this invention. Further, it should be noted that such binding domains could be incorporated at any appropriate location within the polypeptide molecules of the invention.
- [0067] Further embodiments of the invention are thus illustrated by a DNA of the invention further comprising a desired restriction endonuclease site at a desired location and by a polypeptide of the invention further comprising a desired protease cleavage site at a desired location.
  - [0068] The restriction endonuclease site may be introduced so as to facilitate further manipulation of the DNA in manufacture of an expression vector for expressing a polypeptide of the invention; it may be introduced as a consequence of a previous step in manufacture of the DNA; it may be introduced by way of modification by insertion, substitution or deletion of a known sequence. The consequence of modification of the DNA may be that the amino acid sequence is unchanged, or may be that the amino acid sequence is changed, for example resulting in introduction of a desired protease cleavage site, either way the polypeptide retains its first and second domains having the properties required by the invention.
  - [0069] Figure 10 is a diagrammatic representation of an expression product exemplifying features described in this example. Specifically, it illustrates a single polypeptide incorporating a domain equivalent to the light chain of botulinum neurotoxin type A and a domain equivalent to the H<sub>N</sub> domain of the heavy chain of botulinum neurotoxin type A with a N-terminal extension providing an affinity purification domain, namely GST, and a C-terminal extension providing a ligand binding domain, namely an IgG binding domain. The domains of the polypeptide are spatially separated by specific protease cleavage sites enabling selective enzymatic separation of domains as exemplified in the Figure. This concept is more specifically depicted in Figure 11 where the various protease sensitivities are defined for the purpose of example.

#### Assay of product activity

[0070] The LC of botulinum neurotoxin type A exerts a zinc-dependent endopeptidase activity on the synaptic vesicle associated protein SNAP-25 which it cleaves in a specific manner at a single peptide bond. The  $_2$ LH $_{423}$ /A ( $Q_2$ E,N $_{26}$ K, A $_{27}$ Y) (SEQ ID NO: 6) cleaves a synthetic SNAP-25 substrate *in vitro* under the same conditions as the native toxin (Figure 3). Thus, the modification of the polypeptide sequence of  $_2$ LH $_{423}$ /A ( $Q_2$ E,N $_{26}$ K,A $_{27}$ Y) relative to the native sequence and within the minimal functional LC domains does not prevent the functional activity of the LC domains. [0071] This activity is dependent on proteolytic modification of the recombinant GST- $_2$ LH $_{423}$ /A ( $Q_2$ E,N $_{26}$ K,A $_{27}$ Y) to convert the single chain polypeptide product to a disulphide linked dichain species. This is currently done using the proteolytic enzyme trypsin. The recombinant product (100-600 µg/ml) is incubated at 37°C for 10-50 minutes with trypsin (10 µg/ml) in a solution containing 140 mM NaCl, 2.7 mM KCl, 10 mM Na $_2$ HPO $_4$ , 1.8 mM KH $_2$ PO $_4$ , pH 7.3. The

proteolytic enzyme trypsin. The recombinant product (100-600 μg/ml) is incubated at 37°C for 10-50 minutes with trypsin (10 μg/ml) in a solution containing 140 mM NaCl, 2.7 mM KCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 1.8 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.3. The reaction is terminated by addition of a 100-fold molar excess of trypsin inhibitor. The activation by trypsin generates a disulphide linked dichain species as determined by polyacrylamide gel electrophoresis and immunoblotting analysis using polyclonal anti-botulinum neurotoxin type A antiserum.

[0072] <sub>2</sub>LH<sub>423</sub>/A is more stable in the presence of trypsin and more active in the *in vitro* peptide cleavage assay than is <sub>23</sub>LH<sub>423</sub>/A. Both variants, however, are fully functional in the *in vitro* peptide cleavage assay. This demonstrates that the recombinant molecule will tolerate N-terminal amino acid extensions and this may be expanded to other chemical or organic moleties as would be obvious to those skilled in the art.

### 20 Example 2

[0073] As a further exemplification of this invention a number of gene sequences have been assembled coding for polypeptides corresponding to the entire light-chain and varying numbers of residues from the amino terminal end of the heavy chain of botulinum neurotoxin type B. In this exemplification of the disclosure the gene sequences assembled were obtained from a combination of chromosomal and polymerase-chain-reaction generated DNA, and therefore have the nucleotide sequence of the equivalent regions of the natural genes, thus exemplifying the principle that the substance of this disclosure can be based upon natural as well as a synthetic gene sequences.

[0074] The gene sequences relating to this example were all assembled and expressed using methodologies as detailed in Sambrook J, Fritsch E F & Maniatis T (1989) Molecular Cloning: A Laboratory Manual (2nd Edition), Ford N, Nolan C, Ferguson M & Ockler M (eds), Cold Spring Harbor Laboratory Press, New York, and known to those skilled in the art.

[0075] A gene has been assembled coding for a polypeptide of 1171 amino acids corresponding to the entire light-chain (443 amino acids) and 728 residues from the amino terminus of the heavy chain of neurotoxin type B. Expression of this gene produces a polypeptide, LH<sub>728</sub>/B (SEQ ID NO: 20), which lacks the specific neuronal binding activity of full length BoNT/B.

[0076] A gene has also been assembled coding for a variant polypeptide, LH<sub>417</sub>/B (SEQ ID NO: 22), which possesses an amino acid sequence at its carboxy terminus equivalent by amino acid homology to that at the carboxy-terminus of the heavy chain fragment in native LH<sub>N</sub>/A.

[0077] A gene has also been assembled coding for a variant polypeptide, LH<sub>107</sub>/B (SEQ ID NO: 24), which expresses at its carboxy-terminus a short sequence from the amino terminus of the heavy chain of BoNT/B sufficient to maintain solubility of the expressed polypeptide.

#### **Construct Variants**

[0078] A variant of the coding sequence for the first 274 bases of the gene shown in SEQ ID NO: 21 has been produced which whilst being a non-native nucleotide sequence still codes for the native polypeptide.

**[0079]** Two double stranded, a 268 base pair and a 951 base pair, gene sequences have been created using an overlapping primer PCR strategy. The nucleotide bias of these sequences was designed to have an *E.coli* codon usage bias.

[0080] For the first sequence, six oligonucleotides representing the first (5') 268 nucleotides of the native sequence for botulinum toxin type B were synthesised. For the second sequence 23 oligonucleotides representing internal sequence nucleotides 691-1641 of the native sequence for botulinum toxin type B were synthesised. The oligonucleotides ranged from 57-73 nucleotides in length. Overlapping regions, 17-20 nucleotides, were designed to give melting temperatures in the range 52-56°C. in addition, terminal restriction endonuclease sites of the synthetic products were constructed to facilitate insertion of these products into the exact corresponding region of the native sequence. The 268 bp 5' synthetic sequence has been incorporated into the gene shown in SEQ ID NO: 21 in place of the original first 268 bases (and is shown in SEQ ID NO: 27).

Similarly the sequence could be inserted into other genes of the examples.

[0081] Another variant sequence equivalent to nucleotides 691 to 1641 of SEQ ID NO: 21, and employing nonnative codon usage whilst coding for a native polypeptide sequence, has been constructed using the internal synthetic sequence. This sequence (SEQ ID NO: 28) can be incorporated, alone or in combination with other variant sequences, in place of the equivalent coding sequence in any of the genes of the example.

#### Example 3

[0082] An exemplification of the utility of this invention is as a non-toxic and effective immunogen. The non-toxic nature of the recombinant, single chain material was demonstrated by intraperitoneal administration in mice of GST-2LH<sub>423</sub>/A. The polypeptide was prepared and purified as described above. The amount of immunoreactive material in the final preparation was determined by enzyme linked immunosorbent assay (ELISA) using a monoclonal antibody (BA 11) reactive against a conformation dependent epitope on the native LH<sub>N</sub>/A. The recombinant material was serially diluted in phosphate buffered saline (PBS; NaCl 8 g/l, KCl 0.2 g/l, Na<sub>2</sub>HPO<sub>4</sub> 1.15 g/l, KH<sub>2</sub>PO<sub>4</sub> 0.2 g/l, pH 7.4) and 0.5 ml volumes injected into 3 groups of 4 mice such that each group of mice received 10, 5 and 1 micrograms of material respectively. Mice were observed for 4 days and no deaths were seen.

[0083] For immunisation, 20 μg of GST-<sub>2</sub>LH<sub>423</sub>/A in a 1.0 ml volume of water-in-oil emulsion (1:1 vol:vol) using Freund's complete (primary injections only) or Freund's incomplete adjuvant was administered into guinea pigs via two sub-cutaneous dorsal injections. Three injections at 10 day intervals were given (day 1, day 10 and day 20) and antiserum collected on day 30. The antisera were shown by ELISA to be immunoreactive against native botulinum neurotoxin type A and to its derivative LH<sub>N</sub>/A. Antisera which were botulinum neurotoxin reactive at a dilution of 1:2000 were used for evaluation of neutralising efficacy in mice. For neutralisation assays 0.1 ml of antiserum was diluted into 2.5 ml of gelatine phosphate buffer (GPB; Na<sub>2</sub>HPO<sub>4</sub> anhydrous 10 g/l, gelatin (Difco) 2 g/l, pH 6.5-6.6) containing a dilution range from 0.5 μg (5X10-6 g) to 5 picograms (5X10-12 g). Aliquots of 0.5 ml were injected into mice intraperitoneally and deaths recorded over a 4 day period. The results are shown in Table 1 and Table 2. It can clearly be seen that 0.5 ml of 1:40 diluted anti- GST-<sub>2</sub>LH<sub>423</sub>/A antiserum can protect mice against intraperitoneal challenge with botulinum neurotoxin in the range 5 pg - 50 ng (1 - 10,000 mouse LD50; 1 mouse LD50 = 5 pg).

TABLE 1.

Neutralisatio	n of botulir	num neuroto	xin in mice b	y guinea p	oig anti-GS1	- <sub>2</sub> LH <sub>423</sub> /	A antiserum.
		Вс	tulinum Tox	in/mouse	)		
Survivors On Day	<b>0.5μg</b>	<b>0.005μg</b>	0.0005pg	0.5ng	0.005ng	5pg	Control (no toxin)
1	0	4	4	4	4	4	4
2	-	4	4	4	4	4	4
3	-	4	4	4	4	4	4
4	-	4	4	4	4	4	4

TABLE 2.

Neutralisa	tion of bot	tulinum neur	otoxin in mic	e by non-	mmune gui	nea pig	antiserum.				
Botulinum Toxin/mouse											
Survivors On Day	<b>0.5μg</b>	0.005μg	<b>0.0005μg</b>	0.5ng	0.005ng	5pg	Control (no toxin)				
1	0	0	0	0	0	2	4				
2	-	-	-	-	-	0	4				
3	-	-	-	-	-	-	4				
4	-	-	-	-	-	-	4				

### Example 4

30

35

40

45

50

55

Expression of recombinant LH<sub>107</sub>/B in E. coli.

[0084] As an exemplification of the expression of a nucleic acid coding for a  $LH_N$  of a clostridial neurotoxin of a serotype other than botulinum neurotoxin type A, the nucleic acid sequence (SEQ ID NO: 23) coding for the polypeptide

 $LH_{107}/B$  (SEQ ID NO: 24) was inserted into the commercially available plasmid pET28a (Novogen, Madison, WI, USA). The nucleic acid was expressed in *E. coli* BL21 (DE3) (New England BioLabs, Beverley, MA, USA) as a fusion protein with a N-terminal T7 fusion peptide, under IPTG induction at 1 mM for 90 minutes at 37°C. Cultures were harvested and recombinant protein extracted as described previously for  $LH_{423}/A$ .

[0085] Recombinant protein was recovered and purified from bacterial paste lysates by immunoaffinity adsorption to an immobilised anti-T7 peptide monoclonal antibody using a T7 tag purification kit (New England bioLabs, Beverley, MA, USA). Purified recombinant protein was analysed by gradient (4-20%) denaturing SDS-polyacrylamide gel electrophoresis (Novex, San Diego, CA, USA) and western blotting using polyclonal anti-botulinum neurotoxin type antiserum or anti-T7 antiserum. Western blotting reagents were from Novex, immunostained proteins were visualised using the Enhanced Chemi-Luminescence system (ECL) from Amersham. The expression of an anti-T7 antibody and anti-botulinum neurotoxin type B antiserum reactive recombinant product is demonstrated in Figure 13.

[0086] The recombinant product was soluble and retained that part of the light chain responsible for endopeptidase activity.

[0087] The invention thus provides recombinant polypeptides useful inter alia as immunogens, enzyme standards and components for synthesis of molecules as described in WO-A-94/21300.

### SEQUENCE LISTING

#### [0088]

20

(1) GENERAL INFORMATION:

### (i) APPLICANT:

25	<ul> <li>(A) NAME: MICROBIOLOGICAL RESEARCH AUTHORITY</li> <li>(B) STREET: Centre For Applied Microbiology And Research, Porton Down</li> <li>(C) CITY: Salisbury</li> <li>(D) STATE: Wiltshire</li> <li>(E) COUNTRY: UK</li> </ul>
30	(F) POSTAL CODE (ZIP): SP4 OJG
35	(A) NAME: THE SPEYWOOD LABORATORY LIMITED (B) STREET: 14 Kensington Square (C) CITY: London (E) COUNTRY: UK
	(F) POSTAL CODE (ZIP): W8 5HH
	(A) NAME: FOSTER; Keith Alan
40	(B) STREET: Centre For Applied Microbiology And Research, Porton Down (C) CITY: Salisbury
	(D) STATE: Wiltshire
	(E) COUNTRY: UK (F) POSTAL CODE (ZIP): SP4 0JG
	(1) 1 00 ME 00 DE (211). 31 4 000
45	(A) NAME: QUINN; Conrad Padraig
	(B) STREET: Centre For Applied Microbiology And Research, Porton Down
	(C) CITY: Salisbury (D) STATE: Wiltshire
	(E) COUNTRY: UK
50	(F) POSTAL CODE (ZIP): SP4 0JG
	(A) NAME: SHONE; Clifford Charles
	(B) STREET: Centre For Applied Microbiology And Research, Porton Down (C) CITY: Salisbury
55	(D) STATE: Wiltshire
	(E) COUNTRY: UK
	(F) POSTAL CODE (ZIP): SP4 0JG

	(ii) TITLE OF INVENTION: Recombinant Toxin Fragments
	(iii) NUMBER OF SEQUENCES: 28
5	(iv) COMPUTER READABLE FORM:
10	<ul> <li>(A) MEDIUM TYPE: Floppy disk</li> <li>(B) COMPUTER: IBM PC compatible</li> <li>(C) OPERATING SYSTEM: PC-DOS/MS-DOS</li> <li>(D) SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)</li> </ul>
	(2) INFORMATION FOR SEQ ID NO: 1:
15	(i) SEQUENCE CHARACTERISTICS:
	<ul><li>(A) LENGTH: 2616 base pairs</li><li>(B) TYPE: nucleic acid</li><li>(C) STRANDEDNESS: single</li><li>(D) TOPOLOGY: linear</li></ul>
20	(ii) MOLECULE TYPE: DNA (genomic)
	(ix) FEATURE:
25	(A) NAME/KEY: CDS (B) LOCATION:12616
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:
30	
35	
40	
45	
50	
50	

	ATO Met	Glr	TTO	C GTG	L Ası	AAC Lys	G CAG	TTO Phe	C AAG Asi	TAT	: Lys	GA( Asp	C CC	r gr	A AA l As: l:	C GCT n Gly	48
5	GTT Val	GAC Asp	TATI	GCC Ala 20	а Туз	ATC	AAA Lys	ATT	CCA Pro 25	Asr.	GCC Ala	GGG	CAC Glr	S ATO Net	Glı	G CCG	96
10	GTG Val	Lys	GCT Ala 35	Phe	Lys	ATT Ile	CAT His	AAC Asn 40	Lys	ATC Ile	TGG	GTT Val	ATT Ile	Pro	GAA Glu	CGC Arg	144
45	Asp	Thr 50	Phe	Thr	Asn	Pro	Glu 55	Glu	Gly	Asp	Leu	Asn 60	Pro	Pro	Pro	GAA Glu	192
15	Ala 65	Lys	Gln	Val	Pro	Val 70	Ser	Tyr	Tyr	Asp	Ser 75	Thr	Tyr	Leu	Ser	80	240
20	Asp	Asn	Glu	Lys	Asp 85	Asn	Tyr	Leu	Lys	GGA Gly 90	Val	Thr	Lys	Leu	Phe 95	Glu	288
25	Arg	Ile	Tyr	Ser 100	Thr	Asp	Leu	Gly	Arg 105	ATG Met	Leu	Leu	Thr	Ser 110	Ile	Val	336
23	Arg	Gly	Ile 115	Pro	Phe	Trp	Gly	Gly 120	Ser	ACC Thr	Ile	Asp	Thr 125	Glu	Leu	Lys	384
30	Val	Ile 130	Asp	Thr	Asn	Cys	Ile 135	Asn	Val	ATC Ile	Gln	Pro 140	Asp	Gly	Ser	TYT	432
35	Arg 145	Ser	Glu	Glu	Leu	Asn 150	Leu	Val	Ile	ATC Ile	Gly 155	Pro	Ser	Ala	Asp	Ile 160	480
35	Ile	Gln	Phe	Glu	Cys 165	Lys	Ser	Phe	Gly	CAC His 170	Glu	Val	Leu	Asn	Leu 175	Thr	528
40	CGT .	Asn	Gly	Tyr 180	Gly	Ser	Thr	Gln	Tyr 185	Ile	Arg	Phe	Ser	Pro 190	Asp	Phe	576
45	ACG Thr	Phe	Gly 195	Phe	Glu	Glu	Ser	Leu 200	Glu	Val .	Asp '	Thr .	Asn 205	Pro	Leu	Leu	624
45		Ala 210	Gly	Lys	Phe .	Ala	Thr . 215	Asp	Pro .	Ala '	Val :	Thr :	Leu .	Ala :	His (	Glu	672
50	Leu : 225				Gly :					Gly :					Pro .		720

	CGG	C GTO	TTC L Phe	AAC Lys	G GT1 5 Val 245	. Asr	ACC Thr	AAC Asr	GCC 1 Ala	TAC Ty: 250	Ty:	GAC Glu	ATO	AG: Se:	r GG r Gl; 25	T TTA y Leu 5	768
5	GA/ Glu	GTA 1 Val	A AGO	TTC Phe 260	Glu	GAA Glu	CTG Leu	CGC Arg	265	Phe	GG1 Gly	r GGC / Gly	CAT His	GAT Asp 270	Al:	AAG A Lys	816
10	TT7 Phe	T ATC	GAC Asp 275	Ser	TTC Leu	CAG Gln	GAG Glu	AAC Asn 280	Glu	TTC Phe	CG1	CTC Leu	TAC Tyr 285	Ty	TAC Tyl	AAC Asn	864
	AAG Lys	Phe 290	Lys	GAT Asp	ATT	GCA Ala	AGT Ser 295	Thr	CTG Leu	AAC Asn	Lys	GCT Ala 300	Lys	Ser	ATT Ile	GTG Val	912
15	GGT Gly 305	Thr	ACT	GCT Ala	TCA Ser	TTA Leu 310	Gln	TAT	ATG Met	AAA Lys	AAT Asn 315	Val	TTT	AAA Lys	GAG Glu	Lys 320	960
20											Phe					TTA Leu	1008
	Lys	Phe	Asp	Lys 340	Leu	Tyr	Lys	Met	Leu 345	Thr	Glu	Ile	Tyr	Thr 350	Glu	GAT Asp	1056
25	Asn	Phe	Val 355	Lys	Phe	Phe	Lys	Val 360	Leu	Asn	Arg	Lys	Thr 365	Tyr	Leu	AAT Asn	1104
30	Phe	Asp 370	Lys	Ala	Val	Phe	Lys 375	Ile	Asn	Ile	Val	Pro 380	Lys	Val	Asn	Tyr	1152
	Thr 385	Ile	Tyr	Asp	Gly	Phe 390	AAT Asn	Leu	Arg	Asn	Thr 395	Asn	Leu	Ala	Ala	Asn 400	1200
35	Phe	Asn	Gly	Gln	Asn 405	Thr	GAA Glu	Ile	Asn	Asn 410	Met	Asn	Phe	Thr	Lys 415	Leu	1248
	ГЛЗ	Asn	Phe	Thr 420	Gly	Leu	TTT Phe ACT	Glu	Phe 425	Tyr	Lys	Leu	Leu	Cys 430	Val	Arg	1296
40		Ile	11e 435	Thr	Ser	Lys	Thr	Lys 440	Ser	Leu	Asp	Lys	Gly 445	Tyr	Asn	Lys	1344
45	Ala	Leu 450	Asn	Asp	Leu	Сув	11e 455	Lys	Val	Asn	Asn	Trp 460	qaA	Leu	Phe	Phe	
	Ser 465	Pro	Ser	Glu	Asp	Asn 470	Phe '	Thr .	Asn .	Asp	Leu 475	Asn	Lys	Gly	Glu	Glu 480	1440
50	ATT	Thr	Ser	Asp	Thr 485	Asn	Ile	Glu .	Ala	Ala 490	Glu (	Glu .	Asn	Ile	Ser 495	Leu	1488
	GAT Asp	TTA . Leu	Ile	CAA Gln 500	CAA Gln	Tyr '	Tyr 1	Leu '	ACC Thr 1	Phe	AAT Asn	Phe	Asp A	AAT Asn 510	GAA Glu	CCT Pro	1536

	GAA AAT ATT TCA ATA GAA AAT CTT TCA AGT GAC ATT ATA GGC CAA TTA Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu 515 520 525	1584
5	GAA CTT ATG CCT AAT ATA GAA AGA TTT CCT AAT GGA AAA AAG TAT GAG Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu 530 535 540	1632
10	TTA GAT AAA TAT ACT ATG TTC CAT TAT CTT CGT GCT CAA GAA TTT GAA Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu 545 550 555	1680
	CAT GGT AAA TCT AGG ATT GCT TTA ACA AAT TCT GTT AAC GAA GCA TTA His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu 565 570 575	1728
15	TTA AAT CCT AGT CGT GTT TAT ACA TTT TTT TCT TCA GAC TAT GTA AAG Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys 580 585	1776
20	AAA GTT AAT AAA GCT ACG GAG GCA GCT ATG TTT TTA GGC TGG GTA GAA Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu 595 600 605	1824
	CAA TTA GTA TAT GAT TTT ACC GAT GAA ACT AGC GAA GTA AGT ACT ACG Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr 610 620	1872
25	GAT AAA ATT GCG GAT ATA ACT ATA ATT ATT CCA TAT ATA GGA CCT GCT Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala 625 630 640	1920
20	TTA AAT ATA GGT AAT ATG TTA TAT AAA GAT GAT	1968
30	ATA TIT TCA GGA GCT GTT ATT CTG TTA GAA TIT ATA CCA GAG ATT GCA Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala 660 665 670	2016
35	675 680 685	2064
	690 695 Ash Ala Leu Ser Lys Arg Ash Glu	2112
40	705 710 715 The Val The Ash Trp Leu Ala Lys	2160
45	725 730 Het Lys Glu Ala Leu 725 730 735	2208
	740 745 750 750	256
50	755 760 765 765	304
	TTA AGT TCG AAA CTT AAT GAG TCT ATA AAT AAA GCT ATG ATT AAT ATA Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile 770 780	352

5	AA] Asr 785	, rla	TTT Phe	TTG Leu	AAT Asn	CAA Gln 790	Cys	TCT	GTT Val	TCA Ser	TAT Tyr 795	Leu	ATG Met	AAT Asn	TCT Ser	ATG Met 800	2400
v	ATC Ile	Pro	TAT Tyr	GGT	GTT Val 805	AAA Lys	CGG	TTA Leu	GAA Glu	GAT Asp 810	TTT Phe	GAT Asp	GCT Ala	AGT Ser	CTT Leu 815	AAA Lys	2448
10	GAT Asp	GCA Ala	TTA Leu	TTA Leu 820	AAG Lys	TAT Tyr	ATA Ile	TAT Tyr	GAT Asp 825	AAT Asn	AGA Arg	GGA Gly	ACT Thr	TTA Leu 830	ATT Ile	GGT Gly	2496
15	CAA Gln	GTA Val	GAT Asp 835	AGA Arg	TTA Leu	AAA Lys	GAT Asp	AAA Lys 840	GTT Val	AAT Asn	AAT Asn	ACA Thr	CTT Leu 845	AGT Ser	ACA Thr	GAT Asp	2544
15	ATA Ile	CCT Pro 850	TTT Phe	CAG Gln	CTT <b>Le</b> u	TCC Ser	AAA Lys 855	TAC Tyr	GTA Val	GAT Asp	AAT Asn	CAA Gln 860	AGA Arg	TTA Leu	TTA Leu	TCT Ser	2592
20				GAA Glu	Tyr			TAA *									2616
	(2) IN	FORM	MATIO	N FOF	RSEC	ID N	O: 2:										
25	(i	) SEQ	UENC	E CH	ARAC	TERI	STICS	}:									

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 872 amino acids (B) TYPE: amino acid (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein

30

35

40

45

50

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

	Met 1	Gln	Phe	val	Asn 5	Lys	Gln	Phe	Asr	Tyr 10	Lys	Asp	Pro	Va]	Asr 19	
5	Val	Asp	Ile	Ala 20	Tyr	Ile	Lys	Ile	Pro 25	Asn	Ala	Gly	Glm	Met 30		Pro
	Val	Lys	Ala 35	Phe	Lys	Ile	His	Asn 40	Lys	Ile	Trp	Val	Ile 45		Glu	Arg
10	Авр	Thr 50	Phe	Thr	Asn	Pro	Glu 55	Glu	Gly	Ąsp	Leu	Asn 60	Pro	Pro	Pro	Glu
	Ala 65	Lys	Gln	Val	Pro	Val 70	Ser	Tyr	Tyr	Asp	Ser 75	Thr	Tyr	Leu	Ser	Thr 80
15	Asp	Asn	Glu	Lys	Asp 85	Asn	Tyr	Leu	Lys	Gly 90	Val	Thr	Lys	Leu	Phe 95	Glu
	Arg	Ile	Tyr	Ser 100	Thr	Asp	Leu	Gly	Arg 105	Met	Leu	Leu	Thr	Ser 110	Ile	Val
20	Arg	Gly	Ile 115	Pro	Phe	Trp	Gly	Gly 120	Ser	Thr	Ile	Asp	Thr 125	Glu	Leu	Lys
	Val	Ile 130	qsA	Thr	Asn	Сув	Ile 135	Asn	Val	Ile	Gln	Pro 140	Asp	Gly	Ser	Tyr
25	Arg 145	Ser	Glu	Glu	Leu	Asn 150	Leu	Val	lle	Ile	Gly 155	Pro	Ser	Ala	Asp	Ile 160
	Ile	Gln	Phe	Glu	Cys 165	Lys	Ser	Phe	Gly	His 170	Glu	Val	Leu		Leu 175	Thr

	Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp P 180 185 190	'he
5	Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu L 195 200 205	eu
	Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His G 210 215 220	lu
10	Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro As 225 230 235	sn 10
	Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Le 245 250 255	ıu
15	Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Ly 260 265 270	'\$
	Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr As 275 280 285	n
20	Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Va 290 295 300	1
	Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Ly: 305 310 315	
25	Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 330 335	
	Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340 345 350	
30	Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355 360 365	
	Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 380	
35	Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 385 390 395 400	
	Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu . 405 410 415	
40	Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 420 430	
	Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys 435 440 445	
45	Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe 450 460	
	Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu 465 470 480	
50	Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu 485 490 495	
	Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro 500 510	
55	Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu 515 520 525	

	Glu	Leu 530		Pro	Asn	Ile	Glu 535		g Phe	e Pro	Asr	1 Gly 540		s Ly:	s Ty:	r Glu
5	Leu 545		Lys	Tyr	Thr	Met 550		e His	туг	. Leu	Arc 555		Glr	n Gli	u Phe	Glu 560
	His	Gly	Lys	Ser	Arg 565		Ala	Let	Thr	570		· Val	. Ası	ı Glı	1 Ala 579	Leu ;
10	Leu	Asn	Pro	Ser 580	Arg	Val	Tyr	Thr	Phe 585		Ser	Ser	Asp	Ty:		. Lys
	Lys	Val	Asn 595		Ala	Thr	Glu	Ala 600		Met	Phe	Leu	Gly 605		Va]	Glu
15	Gln	Leu 610	Val	Tyr	Asp	Phe	Thr 615		Glu	Thr	Ser	Glu 620		Sez	Thr	Thr
	Asp 625	Lys	Ile	Ala	Asp	Ile 630	Thr	Ile	Ile	Ile	Pro 635	Tyr	Ile	Gly	Pro	Ala 640
20	Leu	Asn	Ile	Gly	Asn 645	Met	Leu	Tyr	Lys	Asp 650	Asp	Phe	Val	Gly	Ala 655	Leu
	Ile	Phe	Ser	Gly 660	Ala	Val	Ile	Leu	Leu 665	Glu	Phe	Ile	Pro	Glu 670		Ala
25	Ile	Pro	Val 675	Leu	Gly	Thr	Phe	Ala 680	Leu	Val	Ser	Tyr	Ile 685	Ala	Asn	Lys
	Val	Leu 690	Thr	Val	Gln	Thr	Ile 695	Asp	Asn	Ala	Leu	Ser 700	Lys	Arg	Asn	Glu
30	Lys 705	Trp	Asp	Glu	Val	Tyr 710	Lys	Tyr	Ile	Val	Thr 715	Asn	Trp	Leu	Ala	Lys 720
					Ile 725	_			_	730	-		-		735	
35	Glu	Asn		Ala 740	Glu	Ala	Thr	Lys	Ala 745	Ile	Ile	Asn	Туг	Gln 750	Tyr	Asn
	Gln		Thr 755	Glu	Glu (	Glu	Lys	760	Asn	Ile	Asn	Phe	Asn 765	Ile	qeA	Asp
40	Leu	Ser 770	Ser	Lys	Leu :		Glu 775	Ser	Ile	Asn		Ala 780	Met	Ile	Asn	Ile
	Asn 1 785	•			•	790					795					800
45	Ile 1	Pro '	Tyr		Val 1 805		Arg			Asp 810		Asp	Ala		Leu 815	Lys
	Asp A	Ala 1		Leu 1 820	Lys :	Tyr	Ile	Tyr	Asp 825	Asn .	Arg	Gly		Leu 830	Ile	Gly
50	Gln 7	1	835	_				840					845			•
	Ile l	Pro 1	Phe (	Gln 1	Leu s		Lys 855	Tyr	Val .	Asp i		Gln . 860	Arg	Leu	Leu	Ser
55	Thr I 865	Phe 1	Chr (	Glu 7		le 1 370	Гуs	*								

(2) INFORMATION FOR SEQ ID NO: 3:

	(i) SEQUENCE CHARACTERISTICS:
5	<ul><li>(A) LENGTH: 2685 base pairs</li><li>(B) TYPE: nucleic acid</li><li>(C) STRANDEDNESS: double</li><li>(D) TOPOLOGY: linear</li></ul>
	(ii) MOLECULE TYPE: DNA (genomic)
10	(ix) FEATURE:
	(A) NAME/KEY: CDS (B) LOCATION:12685
15	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:
20	
25	
30	
35	
40	
45	
50	

	GG G1	A TC y Se 1	C CC r Pr	A GG/ o Gly	A ATT	CAI His	ATC	ACC Thi	TCC r Se	G ACC	r Arg	r CT g Le	G CA u Gl	G AA n Ly	G CT s Le 1	T CTA u Leu 5	48
5	GA. Gl:	A TTO	G GAG	CTC Lev 20	Pro	GGT Gly	ACC Thr	: ATC	G GAC Glu 25	. Phe	GTC Val	AA(	C AA	G CA S G1:	n Ph	C AAC e Asn	96
10	171	r ry:	35 35	Pro	) Val	ASD	GIÀ	40	Asp	Ile	: Ala	Ty-1	11e	Ly:	s Ile	CCA Pro	144
	AAC Lys	TAC Tyz 50	GIY	CAG Gln	ATG Met	CAG Gln	CCG Pro 55	GTG Val	AAG Lys	GCT Ala	TTC Phe	Lys 60	: Ile	CAT His	AA(	AAA Lys	192
15	11e 65	Trp	Val	Ile	CCG Pro	70	Arg	Asp	Thr	Phe	Thr 75	Asn	Pro	Glu	Glu	Gly 80	240
20	Asp	ren	Asn	Pro	CCG Pro 85	Pro	Glu	Ala	Lys	Gln 90	Val	Pro	Val	Ser	Ту <del>г</del> 95	Tyr	288
	Asp	Ser	Thr	100	CTG Leu	Ser	Thr	Asp	Asn 105	Glu	Lys	Asp	Asn	Tyr 110	Leu	Lys	336
25	GIA	Val	Thr 115	Lys	TTA Leu	Phe	Glu	Arg 120	Ile	Tyr	Ser	Thr	Asp 125	Leu	Gly	Arg	384
30	ATG Met	CTG Leu 130	CTG Leu	ACC Thr	TCA Ser	Ile	GTC Val 135	CGC	GGA Gly	ATC Ile	CCA Pro	TTT Phe 140	TGG Trp	GGT Gly	GGC Gly	AGT Ser	432
	ACC Thr 145	ATT Ile	GAE Asp	ACG Thr	GAG :	TTG . Leu : 150	AAG Lys	GTT Val	ATT Ile	Asp	ACT Thr 155	AAC Asn	TGC Cys	ATT Ile	AAC Asn	GTG Val 160	480
35	ATC Ile	CAA Gln	CCA Pro	Asp	GGT A Gly S 165	AGC :	rac Iyr	AGA ' Arg	Ser	GAA Glu 170	GAA Glu	CTT Leu	AAC Asn	CTC Leu	GTA Val 175	ATC Ile	528
40	ATC Ile	GGG Gly	Pro	TCC ( Ser i	GCG ( Ala <i>i</i>	SAC A	ATT :	Ile (	CAG ( Gln ( 185	TTT ( Phe (	GAG ( Glu (	TGC . Cys	Lys	AGC Ser 190	TTT Phe	GGC Gly	576
	CAC His	Glu	GTG Val 195	TTG I	AAC C	TG A	nr 1	CGT 1 Arg 1 200	AAC ( Asn (	GGT :	TAC ( Tyr (	3ly	CT Ser	ACT Thr	CAG '	TAC Tyr	624

	AT Il	T CG e Arg 210	g Phe	C AGO	CCI Pro	A GAC	TTC Phe 215	: Thi	TTO Pho	C GG e Gl	T TT y Ph	C GA e G1 22	u Gl	G AG u Se	C CT	G GAG u Glu	672
5	GT Va. 22	l Asp	Thr	AAC Asr	CCC Pro	CTG Leu 230	Leu	GG1 Gly	GC/ Ala	A GGG	C AA( y Ly: 23:	s Ph	C GC. e Al	A AC	T GAT	r CCA Pro 240	720
10	GC( Ala	GTC a Val	ACC Thr	CTG Lev	GCA Ala 245	His	GAG Glu	CTC	ATC Ile	CAC His 250	Ala	C GG! a Gly	r ca / Hi:	r CG	CTC Lev 255	TAT Tyr	768
	G1)	ATT	GCG Ala	ATT Ile 260	neA	CCG Pro	AAC Asn	CGC	Val 265	Phe	AAC Lys	GTT Val	AA( LASI	270	Asn	GCC Ala	816
15	Туг	Tyr	Glu 275	Met	Ser	Gly	Leu	Glu 280	Val	. Ser	Phe	e Glu	285	Leu	Arg	ACG Thr	864
20	Phe	Gly 290	Gly	His	GAT Asp	Ala	Lys 295	Phe	Ile	Asp	Ser	300	Gln	Glu	Asn	Glu	912
	Phe 305	Arg	Leu	Tyr	TAC	Tyr 310	Asn	Lys	Phe	Lys	Asp 315	Ile	Ala	Ser	Thr	Leu 320	960
25	Asn	Lys	Ala	Lys	TCC Ser 325	Ile	Val	Gly	Thr	Thr 330	Ala	Ser	Leu	Gln	Tyr 335	Met	1008
30					Lys												1056
					GAT Asp												1104
35	Thr	Glu 370	Ile	Tyr	ACA Thr	Glu	Asp 375	Asn	Phe	Val	Lys	Phe 380	Phe	Lys	Val	Leu	1152
	Asn 385	Arg	Lys	Thr		Leu . 390	Asn	Phe	qaA	Lys	Ala 395	Val	Phe	Lys	Ile	Asn 400	1200
40	Ile	Val	Pro	Lys	GTA Val 405	Asn '	Tyr	Thr	Ile	Tyr 410	qaA	Gly	Phe	Asn	Leu 415	Arg	1248
45	Asn	Thr	Asn	Leu 420	GCA Ala	Ala 2	Asn i	Phe .	Asn 425	Gly	Gln	Asn	Thr	Glu 430	Ile	Asn	1296
	Asn	Met .	Asn 435	Phe '	ACT :	Lys 1	Leu !	Lys . 140	Asn	Phe	Thr	Gly	Leu 445	Phe	Glu	Phe	1344
50	Tyr	Lys 450	Leu	Leu (	TGT ( Cys	Val J	Arg (	ily :	Ile	Ile	Thr	Ser 460	Lys	Thr	Lys :	Ser	1392
55					rac I					Asn.					Lys \		1440

	AA As	T AA n As	T TG	G GA	C TTO p Lev 485	1 LUE	TT:	T AG	T CC	T TC. 0 Se: 49	r Gl	A GA u As	T AA p As	T T	ie Ti	CT AAT Ar Asr 95	1488
5	AS	p Le	1 ASI	500	)	, GIO	GIL	1 116	50!	r Sei	r As	p Th	r As	n Il 51	e G1	VA GCA .u Ala	
10	Ali	a Glt	515	ı Asr	ı Ile	ser	Leu	520	Let	ı Ile	e G1:	n Gl	7 Ty:	r Ty	r Le	A ACC	
	Phe	530	) Phe	: Asp	ASD	Glu	Pro 535	GIU	Asn	lle	Se	540	e Glu	) As	n Le	T TCA u Ser	1632
15	Se: 545	AST	Ile	: Ile	GTA	550	Leu	Glu	Leu	Met	9rc 555	) Asr	ılle	: G1:	ı Ar	A TTT g Phe 560	1680
20	Pro	) Asn	Gly	Lys	565	Tyr	Gīu	Leu	Asp	Lys 570	Tyr	Thr	Met	Phe	579		1728
	Leu	Arg	Ala	580 GIN	GLu	Phe	GIA	His	Gly 585	Lys	Ser	Arg	Ile	Ala 590	Let	A ACA	1776
25	Asn	ser	Val 595	Asn	Glu	Ala	Leu	Leu 600	Asn	Pro	Ser	Arg	Val 605	Tyr	Thr	Phe	1824
30	Phe	Ser 610	Ser	Asp	TAT Tyr TGG	Val	Lys 615	Lys	Val	Asn	Lys	Ala 620	Thr	Glu	Ala	Ala	1872
	Met 625	Phe	Leu	Gly	Trp	Val 630	Glu	Gln	Leu	Val	Tyr 635	Asp	Phe	Thr	Asp	Glu 640	1920
35	Thr	Ser	Glu	Val	Ser 645 GGA	Thr '	Thr	Asp	Lys	Ile 650	Ala	Asp	Ile	Thr	Ile 655	Ile	1968
40	Ile	Pro	Tyr	Ile 660	Gly GGT (	Pro 2	Ala .	Leu	Asn 665	Ile	Gly	Asn	Met	<b>Leu</b> 670	Tyr	Lys	2016
	Asp	Asp	Phe 675	Val	Gly :	Ala 1	Leu	11e 680	Phe	Ser (	Gly	Ala	Val 685	Ile	Leu	Leu	2064
45	Glu	Phe 690	Ile	Pro	Glu : GCG :	Ile /	Ala :	Ile :	Pro '	Val 1	Leu	Gly 700	Thr	Phe	Ala	Leu	2112
	Val 705 GCT	Ser	Tyr :	Ile I	Ala /	Asn I 710	ys I	/al i	Leu 1	Thr '	Val 715	Gln	Thr	Ile	Asp	Asn 720	2160
50	Ala :	Leu	Ser 1	Lys i	Arg 3 725	Asn G	ilu I	ys :	[rp	Asp ( 730	3lu '	Val '	Tyr :	Lys	Tyr 735	Ile	2208
55	Val '	Thr	Asn 7	Trp I	Leu A	la L	ys V	al /	Asn 7	Chr G	iln :	ile i	Asp I	750	Ile .	Arg	2256

5	AA Ly	A AA s Ly	A AT 's Me 75	G AA t Lys	A GAZ	A GC	T TT.	A GAI u Glu 760	A ASI	T CA	A GC	A GA a Gl	A GC u Al 76	a Th	A AA r Ly	G GCT s Ala	230
	AT.	A AT e Il 77	A AA e As O	C TAT	CAC Glr	TAT	AA7 ASI 775	r GTL	TAT	T AC	r GA	G GA u Gl: 780	u Gl	G AA u Ly	A AA S As	T AAT n Asn	2352
10	ATT 11e 785		r TT	TAA T naA s	Ile	GAT Asp 790	vař	TTA Leu	AGT Ser	TCC Ser	795	s Let	L AA:	r GAG	TC:	T ATA	2400
15	AAT Asn	Lys	A GCT	ATG Met	ATT Ile 805	AAT Asn	ATA	AAT Asn	AAA Lys	Phe	Lev	AA1 Asn	CAA Glr	TGC Cys	Ser 815		2448
		-,-	200	Met 820	4911	261	MEC	116	825	ıyr	GLY	' Val	Lys	Arg 830	Leu	Glu	2496
20	GAT Asp	TTT	GAT Asp 835	GCT Ala	AGT Ser	CTT Leu	AAA Lys	GAT Asp 840	GCA Ala	TTA Leu	TTA Leu	AAG Lys	TAT Tyr 845	ATA Ile	TAT Tyr	GAT Asp	2544
	AAT Asn	AGA Arg 850	GGA Gly	ACT Thr	TTA Leu	ATT Ile	GGT Gly 855	CAA Gln	GTA Val	GAT Asp	AGA Arg	TTA Leu 860	AAA Lys	GAT Asp	AAA Lys	GTT Val	2592
25	AAT Asn 865	AAT Asn	ACA Thr	CTT Leu	36T	ACA Thr 870	GAT Asp	ATA Ile	CCT Pro	rne	CAG Gln 875	CTT Leu	TCC Ser	AAA Lys	TAC Tyr	GTA Val 880	2640
30	GAT Asp	AAT Asn	CAA Gln	AGA Arg	TTA Leu 885	TTA Leu :	TCT Ser	ACA Thr	Phe '	ACT Thr 890	GAA Glu	TAT Tyr	ATT Ile	Lys	TAA *		2685

## (2) INFORMATION FOR SEQ ID NO: 4:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 895 amino acids (B) TYPE: amino acid

  - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: protein
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

5**5** 

35

40

45

	Gly 1	Ser	Pro	Gly	Ile 5	His	Met	Thr	Ser	Thr 10	Arg	Leu	Gln	Lys	Leu 15	
5	Glu	Phe	Glu	Leu 20	Pro	Gly	Thr	Met	Glu 25	Phe	Val	Asn	Lys	Gln 30		Asn
	Tyr	Lys	Asp 35	Pro	Val	Asn	Gly	Val 40	Asp	Ile	Ala	Tyr	Ile 45	Lys	Ile	Pro
10	Lys	Tyr 50	Gly	Gln	Met	Gln	Pro 55	Val	Lys	Ala	Phe	Lys 60	Ile	His	Asn	Lys
	Ile 65	Trp	Val	Ile	Pro	Glu 70	Arg	Ąsp	Thr	Phe	Thr 75	Asn	Pro	Glu	Glu	Gly 80
15	Asp	Leu	Asn	Pro	Pro 85	Pro	Glu	Ala	Lys	Gln 90	Val	Pro	Val	Ser	Tyr 95	Tyr
	Asp	Ser	Thr	Tyr 100	Leu	Ser	Thr	Asp	Asn 105	Glu	Lys	Asp	Asn	Ту <del>г</del> 110	Leu	Lys
20																

	Gly Val Thr Lys Leu Phe Glu Arg Ile Tyr Ser Thr Asp Leu Gly 115 120 125	Arg
5	Met Leu Leu Thr Ser Ile Val Arg Gly Ile Pro Phe Trp Gly Gly	Ser
	Thr Ile Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn 145 150 155	Val 160
10	Ile Gln Pro Asp Gly Ser Tyr Arg Ser Glu Glu Leu Asn Leu Val 1 165 170 175	Ile
	Ile Gly Pro Ser Ala Asp Ile Ile Gln Phe Glu Cys Lys Ser Phe G 180 185 190	ily
15	His Glu Val Leu Asn Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln T 195 200 205	уr
	Ile Arg Phe Ser Pro Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu G 210 215 220	lu
20		40
	Ala Val Thr Leu Ala His Glu Leu Ile His Ala Gly His Arg Leu Ty 245 250 255	-
25	Gly Ile Ala Ile Asn Pro Asn Arg Val Phe Lys Val Asn Thr Asn Al 260 270	
	Tyr Tyr Glu Met Ser Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Th 280 285	
30	Phe Gly Gly His Asp Ala Lys Phe Ile Asp Ser Leu Gln Glu Asn Gl 290 295 300	
	Phe Arg Leu Tyr Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Le 305 310 315 32	0
35	Asn Lys Ala Lys Ser Ile Val Gly Thr Thr Ala Ser Leu Gln Tyr Mei 325 330 335	
	Lys Asn Val Phe Lys Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly 340 345 350	-
40	Lys Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu 355 360 365	
	Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu 370 380  Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn	
45	385 390 395 The Lys Ile Asn 395 400  Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg	ì
	Asn Thr Asn Leu Ala Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn	
50	Asn Met Asn Phe Thr Lys Leu Lys Asn Phe Thr Gly Leu Phe Glu Phe	
	435 440 445  Tyr Lys Leu Leu Cys Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser	
55	450 455 460	

	Le:	u Asp 5	) Lys	: Gly	Туг	470	n Ly	s Al	a Le	u As	n As 47	p Le S	u Cy	's I.	le L	ys Val 480
5	Ası	n Asr	Trp	Asp	Leu 485	Phe	e Pho	e Se	r Pr	o Se:	r Gl	u As	p As	n Pl		nr Asn 95
	Asp	Leu	. Asn	Lys 500	Gly	Gli	Gl:	ı Il	2 Th: 50!	r Se:	r As <sub>l</sub>	p Th	r As	n I]		u Ala
10	Ala	Glu	Glu 515	Asn	Ile	Ser	Let	3 Asy 520	Let	ı Ile	e Gla	n Gl	n Ty 52		r Le	u Thr
	Phe	Asn 530	Phe	Asp	Asn	Glu	9rc 535	Glu	AST	ılle	: Se	11e 540		u As	n Le	u Ser
15	Ser 545	Asp	Ile	Ile	Gly	Gln S50	Leu	Glu	Leu	Met	Pro 555	Asr	110	e Gl	u Ar	g Phe 560
					565					570					57	
20				580					585					59	)	u Thr
			595					600				•	605			Phe Phe
25		610					615					620				Ala
	625			Gly		630					635				_	640
30					645				,	650					655	
			'	Ile ( 660 Val (					665					670	-	_
35	Glu		675					680					685			
	1	690		•			695					700				
40	Val . 705				7	/10				•	715				_	720
	Ala 1			7	25					730					735	
45	Val 7		7	40					/45					750		
	Lys I	7	755					760				•	765			
50		70				7	775				•	780		_		
	Ile A 785				7	90				7	95					800
55	Asn L	ys A	la M	et I:	le A 05	sn I	le A	ksn I	ys E	Phe L	eu A	lsn (	iln (		Ser Ser	Val

								_		333 0	10 0	ı				
	Ser	Tyr	Leu	Met 820	Asn	Ser	Met	Ile	Pro 825	Tyr	Gly	Val	Lys	Arg 830	Leu	Glu
5	Asp	Phe	<b>Asp</b> 835	Ala	Ser	Leu	Lys	Asp 840	Ala	Leu	Leu	Lys	Tyr 845	Ile	Tyr	Asp
	Asn	Arg 850	Gly	Thr	Leu	Ile	Gly 855	Gln	Val	Asp	Arg	Leu 860	Lys	Asp	Lys	Val
10	Asn 865	Asn	Thr	Leu	Ser	Thr 870	Asp	Ile	Pro	Phe	Gln 875	Leu	Ser	Lys	Tyr	Val 880
	Asp	Asn	Gln	Arg	Leu 865	Leu	Ser	Thr	Phe	Thr 890	Glu	Tyr	Ile	Lys	* 895	
15	(2) ا	NFOF	RMATIO	ON FO	R SE	Q ID N	VO: 5:									
		(i) SE	QUEN	ICE CI	HARA	CTER	ISTIC	S:								
		(A	A) LEN	GTH:	2622	base p	oairs									
20			3) TYP C) STR				uble									
			) TOP				GDIO									
25			OLECL () FEA			NA (g	genom	ic)								
			NAM ) LOC													
30																
00	,	(XI) SE	QUEN	NCE D	ESUR	IPHO	IN: SE	ו עו טי	NO: 5:							
35																
40																

	GGA Gly 1	TCC Ser	Met	GAG Glu	TTC Phe 5	GTG Val	AAC Asn	AAG Lys	CAG Gln	TTC Phe 10	AAC	TAT	Lys	GAC Asp	Pro 15	GTA Val	41
5	AAC Asn	GGT Gly	GTT Val	GAC Asp 20	Ile	GCC Ala	TAC	ATC Ile	AAA Lys 25	ATT	CCA Pro	AAG Lys	TAC	GGC Gly 30	CAG Gln	ATG Met	96
10	CAG Gln	CCG Pro	GTG Val 35	AAG Lys	GCT Ala	TTC	AAG Lys	ATT Ile 40	CAT His	AAC Asn	AAA Lys	ATC Ile	TGG Trp 45	GTT Val	ATT Ile	CCG Pro	144
	GAA Glu	CGC Arg 50	GAT Asp	ACA Thr	TTT Phe	ACG Thr	AAC Asn 55	CCG Pro	GAA Glu	GAA Glu	GGA Gly	GAC Asp 60	TTG Leu	AAC Asn	CCG Pro	CCG Pro	192
15	CCG Pro 65	GAA Glu	GCA Ala	AAG Lys	CAG Gln	GTG Val 70	CCA Pro	GTT Val	TCA Ser	TAC Tyr	TAC Tyr 75	gat Asp	TCA Ser	ACC Thr	TAT Tyr	CTG Leu 80	240
20	AGC Ser	ACA Thr	gac Asp	AAC Asn	GAG Glu 85	AAG Lys	GAT Asp	AAC Asn	TAC Tyr	CTG Leu 90	AAG Lys	GGA Gly	GTG Val	ACC Thr	AAA Lys 95	TTA Leu	288
	TTC Phe	GAG Glu	CGT Arg	ATT Ile 100	TAT Tyr	TCC Ser	ACT Thr	GAC Asp	CTG Leu 105	GGC Gly	CGT Arg	ATG Met	CTG Leu	CTG Leu 110	ACC Thr	TCA Ser	336
25	ATC Ile	GTC Val	CGC Arg 115	GGA Gly	ATC Ile	CCA Pro	TTT Phe	TGG Trp 120	GGT Gly	GGC Gly	AGT Ser	ACC Thr	ATT Ile 125	GAC Asp	ACG . Thr	GAG Glu	384
30	TTG Leu	AAG Lys 130	GTT Val	ATT Ile	GAC Asp	ACT Thr	AAC Asn 135	TGC Cys	ATT Ile	AAC Asn	Val	ATC Ile 140	CAA Gln	CCA Pro	GAC Asp	GGT Gly	432

	AGG Set 145	r Ty	c AGA	A TCT g Sei	GAP	GAA Glu 150	ı Lev	T AAG L ASI	cro n Lev	C GT	A ATO 1 Ile 159	e Il	C GG e Gl	G CC y Pr	C TC o Se	C GCG r Ala 160	480
5	GAC Asp	ATT Ile	r ATC	CAC Glr	TTI Phe 165	Glu	TGC Cys	Lys	3 AGC 5 Ser	Phe 170	: Gly	CA Ri	C GA s Gl	A GT u Va	G TT l Le 17	G AAC u Asn 5	528
10	CTG Leu	ACG Thr	G CGI	AAC Asn 180	Gly	TAC	GGC	TCI Ser	Thr 185	Glr	TAC Tyr	T Il	r cg	T TT g Ph	e Se	C CCA	576
	GAC Asp	TTC Phe	ACG Thr 195	Phe	GGT Gly	TTC Phe	GAG Glu	GAG Glu 200	Ser	CTG	GAG Glu	GT:	GA: L Asi 205	Th	C AAC C Asi	CCG Pro	624
15	CTG Leu	TTG Leu 210	Gly	GCA Ala	GGC Gly	AAG Lys	TTC Phe 215	GCA Ala	ACT Thr	GAT Asp	CCA Pro	GCC Ala 220	. Val	ACC Thi	CTC Lev	GCA Ala	672
20	His 225	Glu	Leu	Ile	His	Ala 230	Gly	His	Arg	Leu	Tyr 235	Gly	Ile	Ala	Ile	AAC Asn 240	720
	Pro	Asn	Arg	Val	TTC Phe 245	Lys	Val	Asn	Thr	Asn 250	Ala	Tyr	Tyr	Glu	Met 255	Ser	768
25	Gly	Leu	Glu	Val 260	AGC Ser	Phe	Glu	Glu	Leu 265	Arg	Thr	Phe	Gly	Gly 270	His	Asp	816
30	Ala	Lys	Phe 275	Ile	GAC Asp	Ser	Leu	Gln 280	Glu	Asn	Glu	Phe	Arg 285	Leu	Тут	Tyr	864
	Tyr	Аsп 290	Ľýs	Phe	AAA Lys	Asp	11e 295	Ala	Ser	Thr	Leu	Asn 300	Lys	Ala	Lys	ser	912
35	Ile 305	Val	Gly	Thr		Ala 310	Ser	Leu	Gln	Tyr	Met 315	Lys	Asn	Val	Phe	Lys 320	960
40	Glu	Lys	Tyr	Leu	CTA Leu 325	Ser	Glu	Asp	Thr	Ser 330	Gly	Lys	Phe	Ser	Val 335	Asp	1008
	Lys	Leu	Lys	Phe 340	GAT Asp	Lys	Leu	Tyr	Lys 345	Met	Leu	Thr	Glu	Ile 350	Tyr	Thr	1056
45	Glu	Asp	Asn 355	Phe	GTT Val	Lys	Phe	Phe 360	Lys	Val	Leu	Asn	Arg 365	Lys	Thr	Tyr	1104
50	Leu	Asn 370	Phe	Asp	AAA Lys	Ala	Val 375	Phe	Lys	Ile	Asn	Ile 380	Val	Pro	Lys	Val	1152
	AAT Asn 385	Tyr	Thr	Ile	Tyr	ABP (	Gly	Phe .	Asn :	Leu .	Arg 395	Asn	Thr	Asn	Leu .	Ala 400	1200
55	GCA Ala			Asn	GGT ( Gly ( 405				Glu :					Asn			1248

_	AA. Ly:	A CT	A AA	A AAT S AST 420	ı Phe	C ACT	GGF Gly	TTC	TTT 1 Pho 425	e Gl	A TT	r TA:	r AA	G TTO	u Le	A TGT u Cys	1296
5	GT/ Va.	A AGA	435 435	/ Ile	ATA	ACT Thr	TCI Ser	Lys 440	Thi	C AAJ C Lys	A TC/ S Sei	A TTA	A GAT 1 Asj 445	Ly:	A GG s Gl	A TAC y Tyr	1344
10	AAT Asr	AAC Lys 450	. Ala	TTA Leu	AAT Asn	GAT Asp	TTA Leu 455	Cys	ATC 11e	AA! Lys	A GTT Val	AAT L Ast 460	Ası	TGC Tr	GA(	C TTG p Leu	1392
15	777 Phe 465	: Phe	AGI Ser	CCT Pro	TCA Ser	GAA Glu 470	GAT Asp	AAT Asn	TTT Phe	ACT Thr	AA1 Asn 475	Asp	CTA Leu	AA1 Asn	C AAJ	A GGA Gly 480	1440
13	GAA Glu	GAA Glu	ATI	ACA Thr	TCT Ser 485	Ąsp	ACT Thr	AAT Asn	ATA Ile	GAA Glu 490	Ala	GCA Ala	GAA Glu	GAA Glu	AA7 Asr 495	ATT Ile	1488
20	Ser	Leu	Asp	Leu 500	Ile	Gln	Gln	Tyr	Tyr 505	Leu	Thr	Phe	Asn	Phe 510	Asp	AAT Asn	1536
05	Glu	Pro	Glu 515	Asn	Ile	Ser	Ile	Glu 520	Asn	Leu	Ser	Ser	Asp 525	Ile	Ile	GGC Gly	1584
25	Gln	Leu 530	Glu	Leu	Met	Pro	Asn 535	Ile	Glu	Arg	Phe	Pro 540	Asn	Gly	Lys	.*	1632
30	Tyr 545	Glu	Leu	Asp	Lys	TAT Tyr 550	Thr	Met	Phe	His	Tyr 555	Leu	Arg	Ala	Gln	Glu 560	1680
	Phe	Glu	His	Gly	Lys 565	Ser	Arg	Ile	Ala	Leu 570	Thr	Asn	Ser	Val	Asn 575	Glu	1728
35	Ala	Leu	Leu	Asn 580	Pro	Ser	Arg	Va1	Tyr 585	Thr	Phe	Phe	Ser	Ser 590	Asp	Tyr	1824
40	Val	Lys	Lys 595	Val	Asn	Lys TAT	Ala	Thr 600	Glu	Ala	Ala	Met	Phe 605	Leu	Gly	Trp	1872
	Val	Glu 610 ACG	Gln GAT	Leu	Val ATT	Tyr .	Asp 615 GAT	Phe ATA	Thr	Asp ATA	Glu	Thr 620 ATT	Ser	Glu TAT	Val ATA	Ser GGA	1920
45	Thr 625	Thr	Asp	Lys aat	Ile .	Ala 630 GGT	Asp AAT .	Ile ATG	Thr TTA	Ile TAT	Ile 635 AAA	Ile GAT	Pro GAT	Tyr	Ile GTA	Gly 640 GGT	1968
50	Pro	Ala TTA	Leu ATA	Asn TTT	Ile 645 TCA	Gly :	Asn GCT	Met GTT	Leu Att	Tyr 650 CTG	Lys	Asp Gaa	Asp TTT	Phe ATA	Val 655 CCA	Gly GAG	2016
	Ala	Leu GCA	Ile ATA	Phe 660 CCT	Ser	Gly i	Ala GGT .	Val ACT	Ile 665 TTT	Leu GCA	Leu	Glu GTA	Phe TCA	Ile 670 TAT	Pro ATT	GCG GCG	2064
55	Ile	Ala	11e 675	Pro	Val	Leu (	Gly '	Thr 680	Phe .	Ala	Leu	Val .	Ser 685	Tyr	Ile	Ala	

	AA? Ası	AA( 2 Lys 690	s Val	CTA Lei	A ACC	GTT Val	CAA Glr 695	l Thi	ATA : Ile	GAT Asp	AA1 Asr	GCT Ala 700	a Lei	A AG	r AA	A AGA s Arg	2112
5	AAI Asn 705	Glu	AAA Lys	Trp	GAT Asp	GAG Glu 710	. Val	TAT	AAA Lys	TAT	ATA Ile 715	· Val	ACA Tha	A AAT Asr	TG(	TTA Leu 720	2160
10	GCA Ala	AAG Lys	GTT Val	AAT Asn	ACA Thr 725	Gln	ATT	GAT Asp	CTA Leu	ATA Ile 730	Arg	AAA Lys	AAA Lys	ATG Met	Lys 735	GAA Glu	2208
	GCT Ala	TTA	GAA Glu	AAT Asn 740	Gln	GCA Ala	GAA Glu	GCA Ala	ACA Thr 745	AAG Lys	GCT Ala	ATA Ile	ATA Ile	AAC Asn 750	Tyr	CAG Gln	2256
15	Tyr	Asn	Gln 755	Tyr	Thr	Glu	Glu	Glu 760	Lys	Asn	Asn	Ile	Asn 765	Phe	Asn		2304
20	Asp	770	TTA Leu	Ser	Ser	Lys	Leu 775	Asn	Glu	Ser	Ile	Asn 780	Lys	Ala	Met	Ile	2352
	Asn 785	Ile	AAT Asn	Lys	Phe	Leu 790	Asn	Gln	Cys	Ser	Val 795	Ser	Tyr	Leu	Met	Asn 800	2400
25	Ser	Met	ATC Ile	Pro	Tyr 805	Gly	Val	Lys	Arg	Leu 810	Glu	Asp	Phe	Asp	Ala 815	Ser	2448
30	Leu	Lys	GAT Asp	Ala 820	Leu	Leu	Lys	Tyr	Jle 825	Tyr	Asp	Asn	Arg	Gly 830	Thr	Leu	2496
	Ile	Gly	CAA Gln 835	Val	Asp	Arg	Leu	Lys 840	Asp	Lys	Val	Asn	Asn 845	Thr	Leu	Ser	2544
35		Asp 850	Ile	Pro	Phe	Gln	Leu 855	Ser	Lys	Tyr	Val .						2592
	TTA Leu 865				Thr					*							2622

40

45

(2) INFORMATION FOR SEQ ID NO: 6:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 874 amino acids (B) TYPE: amino acid (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

55

Gly 1	Ser	Met	Glu	Phe 5	Val	Asn	Lys	Gln	Phe 10	Asn	Tyr	Lys	Asp	Pro 15	Val
Asn	Gly	Val	Asp 20	Ile	Ala	Tyr	Ile	Lys 25	Ile	Pro	Lys	Tyr	Gly 30	Gln	Met
Gln	Pro	Val 35	Lys	Ala	Phe	Lys	Ile 40	His	Asn	Lys	Ile	Trp 45	Val	Ile	Pro

	Glu	Arg 50	Asp	Thr	Phe	Thi	As:	n Pro	o Gl	u Gli	u Gly	As 6		u As	n Pr	o Pro
5	Pro 65	Glu	Ala	Lys	Glr	70	. Pro	o Va	l Se	г Ту	r Ty:	Ası	p Se:	r Th	г Ту	r Leu 80
	Ser	Thr	Asp	Asn	. Glu 85	Lys	Asp	) Ası	туз	Le:		Gly	/ Vai	l Thi	r Ly 9	s Leu 5
10	Phe	Glu	Arg	Ile 100	Tyr	Ser	Thi	Ası	105	Gly	Arg	Met	: Leu	le.		r Ser
	Ile	Val	Arg 115	Gly	Ile	Pro	Phe	120	Gly	Gly	Ser	Thr	11e		Th:	r Glu
15	Leu	Lys 130	Val	Ile	Asp	Thr	Asn 135	Сув	Ile	Asn	Val	11e 140		Pro	Asp	Gly
	Ser 145	Tyr	Arg	Ser	Glu	Glu 150	Leu	Asn	Leu	Val	Ile 155	Ile	Gly	Pro	Sez	Ala 160
20				Gln	165					170					175	,
				Asn 180					185			•		190		
25			195	Phe				200					205			
		210		Ala			215					220				
30	225			Ile		230					235					240
				Val	245					250					255	
35	Gly			260					265					270		<del>-</del>
	Ala :		275	•				280					285			_
40		290					295					300			-	
	Ile '	Val (	31y '	Thr '		Ala 310	Ser	Leu	Gln		Met 315	Lys	Asn	Val	Phe	Lys 320
45	Glu l	Lys :	Tyr 1		Leu : 325	Ser (	Glu	Asp		Ser 330	Gly	Lys-	Phe		Val 335	Asp
	Lys I	Leu 1		Phe 2 340	Asp :	Lys :	Leu	Tyr	Lys   345	Met	Leu '	Thr		Ile ' 350	Tyr	Thr
50	Glu A		Asn 1 855	Phe V	/al :	Lys :		Phe 360	Lys '	Val:	Leu i		Arg : 365	Lys :	Thr	Tyr
	Leu A	Asn I	Phe A	Asp 1	Lys I	Ala :	Val 375	Phe	Lys :	Ile :		lle '	Val 1	Pro I	Lys	Val
55	Asn 1	yr 1	thr 1	lle T		Asp ( 390	Sly	Phe .	Asn :		Arg 1 395	Asn (	Thr i	Asn 1		Ala 400

	Ala	a Ası	n Phe	e Ası	1 Gl 40	y Gli 5	n Ası	n Th:	r Gl	110 410		n As:	n Mei	t As:	n Ph 41	e Thr 5
5	Lys	s Le	l Lys	420	n Pho	e Thi	r Gly	y Let	2 Phe 425	e Glu	ı Phe	Ty:	c Lys	430		u Cys
	Val	Arg	Gly 435	r Ile	: Ile	e Thi	Ser	Lys 440	Thi	: Lys	Ser	Lev	1 Asp 445		Gly	y Tyr
10	Asn	Lys 450		Lev	Ası	a Asp	Lev 455	Cys	Ile	Lys	Val	. Asr 460		Tr	) Asp	Leu
	Phe 465		Ser	Pro	Ser	Glu 470	Asp	) Asn	Phe	Thr	Asn 475		Leu	Asr	Lys	Gly 480
15	Glu	Glu	Ile	Thr	Sez 485	Asp	Thr	Asn	Ile	Glu 490		Ala	Glu	Glu	Asn 495	Ile
	Ser	Leu	Asp	Leu 500	Ile	Gln	Gln	Tyr	Tyr 505		Thr	Phe	Asn	Phe 510	Asp	Asn
20	Glu	Pro	Glu 515	Asn	Ile	Ser	Ile	Glu 520	Asn	Leu	Ser	Ser	Asp 525	Ile	Ile	Gly
	Gln	Leu 530	Glu	Leu	Met	Pro	Asn 535	Ile	Glu	Arg	Phe	Pro 540	Asn	Gly	Lys	Lys
25	Tyr 545	Glu	Leu	Asp	Lys	Tyr 550	Thr	Met	Phe	His	Tyr 555	Leu	Arg	Ala	Gln	Glu 560
	Phe	Glu	His	Gly	Lys 565	Ser	Arg	Ile	Ala	Leu 570	Thr	Asn	Ser	Val	Asn 575	Glu
30	Ala	Leu	Leu	Asn 580	Pro	Ser	Arg	Val	Tyr 585	Thr	Phe	Phe	Ser	Ser 590	Asp	Tyr
	Val	Lys	Lys 595	Val	Asn	Lys	Ala	Thr 600	Glu	Ala	Ala	Met	Phe 605	Leu	Gly	Trp
35	Val	Glu 610	Gln	Leu	Val	Tyr	Asp 615	Phe	Thr	Asp	Glu	Thr 620	Ser	Glu	Val	Ser
	625			•		Ala 630					635					640
40	Pro	Ala	Leu		Ile 645	Gly	Asn	Met		Tyr 650	Lys	Asp	Asp		Val 655	Gly
	Ala	Leu		Phe 660	Ser	Gly	Ala	Val	Ile 665	Leu	Leu	Glu		Ile 670	Pro	Glu
45	Ile .	Ala	Ile 675	Pro	Val	Leu		Thr 680	Phe .	Ala	Leu '		Ser 685	Tyr	Ile	Ala
	Asn	Lys 690	Val	Leu	Thr		Gln <b>69</b> 5	Thr	Ile .	Asp .		Ala 700	Leu :	Ser	Lys .	Arg
50	Asn (	Glu	Lys '	Trp /		Glu 710	Val	Tyr :	Lys '		Ile ' 715	Val '	Thr i	Asn '	-	Leu 720
	Ala 1	ràs .	Val :		Thr 725	Gln	Ile .	Asp :		11e / 730	Arg 1	Lys :	Lys i		Lys (	Glu
55	Ala 1	Leu (		Asn ( 740	3ln .	Ala (	Glu /		Thr 1 745	Lys )	Ala :	(le :		Asn :	Cyr (	Gln

	Туг	: Asn	Gln 755	Туг	Thr	Glu	Glu	Glu 760	Lys	Asn	Asn	Ile	Asn 765	Phe	Asn	Ile
5	Asp	770	Leu	Ser	Ser	Lys	<b>Le</b> u 775	Asn	Glu	Ser	Ile	Asn 780	Lys	Ala	Met	Ile
	Asn 785	Ile	Asn	Lys	Phe	Leu 790	Asn	Gln	Cys	Ser	Val 795	Ser	Tyr	Leu	Met	Asn 800
10	Ser	Met	Ile	Pro	Tyr 805	Gly	Val	Lys	Arg	Leu 810	Ğlu	Asp	Phe	Asp	Ala 815	Ser
	Leu	Lys	Asp	Ala 820	Leu	Leu	Lys	Tyr	Ile 825	Tyr	Asp	Asn	Arg	Gly 830	Thr	Leu
15	Ile	Gly	Gln 835	Val	Asp	Arg	Leu	Lys 840	Asp	Lys	Val	Asn	Asn 845	Thr	Leu	Ser
	Thr	Asp 850	Ile	Pro	Phe	Gln	Leu 855	Ser	Lys	Tyr	Val	Asp 860	Asn	Gln	Arg	Leu
20	Leu 865	Ser	Thr	Phe	Thr	Glu 870	Tyr	Ile	Lys	*						
	(2) II	NFOR	MATIC	ON FO	R SE	י סו ב	NO: 7:							-		
25		(i) SEC	QUEN	CE CI	HARAG	CTER	ISTIC	S:								
			-		2613 l		oairs									
30		(C	) STR	ANDE	eleic ac EDNES SY: line	S: do	uble									
	(	(ii) MO	LECU	LE TY	PE: D	NA (g	genom	ic)								
35	(	ix) FE	ATUR	E:												
,,					Y: CDS I:12											
<b>1</b> 0	(	xi) SE	QUEN	ICE D	ESCR	IPTIO	N: SE	Q ID	NO: 7:							

5	ATC Met	CCA Pro	TTT Phe	Val	AAT Asn 5	AAA Lys	CAA Gln	TTI Phe	AAT Asn	TAT Tyr 10	Lys	GAT Asp	CCT Pro	GTA Val	AAT Asn 15	GGT	48
v	GTT Val	GAT Asp	ATT Ile	GCT Ala 20	TAT	ATA Ile	AAA Lys	ATT	CCA Pro 25	AAT Asn	GCA Ala	GGA Gly	CAA Gln	ATG Met 30	Gln	CCA	96
10	GTA Val	AAA Lys	GCT Ala 35	TTT	AAA Lys	ATT Ile	CAT His	AAT Asn 40	AAA Lys	ATA Ile	TGG Trp	GTT Val	ATT Ile 45	CCA Pro	GAA Glu	AGA Arg	144
15	GAT Asp	ACA Thr 50	TTT Phe	ACA Thr	AAT Asn	CCT Pro	GAA Glu 55	GAA Glu	GGA Gly	GAT Asp	TTA Leu	AAT Asn 60	CCA Pro	CCA Pro	CCA	GAA Glu	192
13	GCA Ala 65	AAA Lys	CAA Gln	GTT Val	CCA Pro	GTT Val 70	TCA Ser	TAT	TAT Tyr	GAT Asp	TCA Ser 75	ACA Thr	TAT Tyr	TTA Leu	AGT Ser	ACA Thr 80	240
20	GAT Asp	AAT Asn	GAA Glu	AAA Lys	GAT Asp 85	AAT Asn	TAT Tyr	TTA Leu	AAG Lys	GGA Gly 90	GTT Val	ACA Thr	AAA Lys	TTA Leu	TTT Phe 95	GAG Glu	288

	AG! Arg	A AT	T TA' e Ty:	T TC	r Th	T GAT r Asp	CT?	r GG/ ı Gl/	A AG	g Me	G TI t Le	G TI	'A AC	A TO r Se 11	r I	CA GTA .e Val	336
5	AGC Arg	G GG	A ATA	e Pro	A TT	r TG0 ∋ Trp	GG1 Gly	GG/ Gl <sub>2</sub> 120	/ Se	r AC	A AT	A GA e As	T AC p Th 12	r Gl	A TI u Le	'A AAA u Lys	384
10	GTT Val	116 130	e Asp	r ACT	TAA 7 Rar	TGT Cys	Ile 135	Asn	GTC Val	AT:	A CA e Gl	A CC n Pr 14	o As	T GG p Gl	T AG y Se	T TAT	432
	Arg 145	Ser	Glu	Glu	Leu	150	Leu	Val	Ile	: Ile	15	y Pr	o Se:	r Al	a As	T ATT P Ile 160	480
15	Ile	Gln	Phe	Glu	Cys 165	Lys	Ser	Phe	Gly	His 170	Glı	ı Va	l Le	ı Ası	1 Le 17		528
20	Arg	Asn	Gly	Tyr 180	Gly	Ser	Thr	Gln	Tyr 185	Ile	Arg	Phe	e Ser	190	As;	r TTT p Phe	576
	Thr	Phe	Gly 195	Phe	Glu	Glu	Ser	Leu 200	Glu	Val	Asp	Thi	205	Pro	Lei	TTA Leu	624
25	Gly	Ala 210	Gly	Lys	Phe	Ala	Thr 215	Asp	Pro	Ala	Val	Thr 220	Leu	Ala	His	GAA Glu	672
30	Leu 225	Ile	His	Ala	Gly	His 230	Arg	Leu	Tyr	Gly	11e 235	Ala	Ile	Asn	Pro	AAT Asn 240	720
	Arg	Val	Phe	Lys	Val 245	Asn	Thr	Asn	Ala	Tyr 250	Tyr	Glu	Met	Ser	Gly 255		768
35	Glu	Val	Ser	Phe 260	Glu	Glu	Leu	Arg	Thr 265	Phe	Gly	Gly	His	Asp 270	Ala	Lys	816 864
40	Phe	Ile	Asp 275	Ser	Leu	Gln	Glu	Asn 280	Glu	Phe	Arg	Leu	Tyr 285	Tyr	Tyr	Asn	912
	Lys	Phe 290	Lys	qaA	Ile	Ala	Ser 295	Thr	Leu	Asn	Lys	Ala 300	Lys	Ser	Ile	Val	
45	Gly 305	Thr	Thr	Ala	Ser	Leu ( 310	3ln '	Tyr 1	Met	Lys	Asn 315	Val	Phe	Lys	Glu	Lys 320	960
50	TAT (	Leu	Leu	Ser	Glu . 325	Asp :	Chr !	Ser (	Gly	Lys 330	Phe	Ser	Val	Asp	Lys 335	Leu	1008
55	Lys 1	Phe .	Asp	Lys : 340	Leu '	Tyr I	Lys I	let 1	Leu ' 345	Thr	Glu	Ile	Tyr	Thr 350	Glu	Asp	1056
5 <b>5</b>	AAT :	?he '					ys \										1104

5	TT: Phe	GA' As;	p Ly	A GC s Al	c gt a Va	A TTT	CAAC Lys 375	: 116	A AA' B As:	r at	A GT e Va	A CC 1 Pr 38	o Ly	G GI s Va	'A AA	TAI	C 1152
5	ACA Thi 385	: Ile	A TA'	r GA' r Asj	r <i>GGI</i> p Gly	777 Phe 390	: Asn	TTA Lev	A AG	A AA' J Asi	T AC. n Th: 39	r As	T TT n Le	A GC u Al	A GC a Al	A AA( a Ası 40(	n
10	Phe	e Ası	n Gly	/ Gl:	1 Asr 405	Thr	Glu	Ile	: Asr	410	n Mei	t Ası	n Ph	e Th	r Ly 41		
15	Lys	Asr	ı Phe	420	r Gly	Leu	Phe	Glu	425	Туз	: Lys	s Leu	ı Lei	1 Cy:	ya D	A AGA l Arg	1
73	Gly	Ile	11e 435	Thr	: Ser	Lys	Thz	Lys 440	Ser	Leu	y yst	Lys	G13	y Tyn	As	r AAG n Lys	
20	Ala	Leu 450	Asn	Asp	Leu	Cys	Ile 455	Lys	Val	Asn	) Asn	460	Asp	Lei	Phe	TTT Phe	
25	Ser 465	Pro	Ser	Glu	Asp	Asn 470	Phe	Thr	Asn	Asp	Leu 475	Asn	Lys	Gly	Gli	GAA Glu 480	1440
25	Ile	Thr	Ser	Asp	Thr 485	Asn	Ile	Glu	Ala	Ala 490	Glu	Glu	Asn	Ile	Ser 495	i	1488
30	Asp	Leu	Ile	Gln 500	Gln	TAT Tyr GAA	Tyr	Leu	Thr 505	Phe	Asn	Phe	Asp	Asn 510	Glu	Pro	1536
	Glu	Asn	Ile 515	Ser	Ile	Glu ATA	aeA	Leu 520	Ser	Ser	Asp	Ile	Ile 525	Gly	Gln	Leu	1584
35	Glu	Leu 530	Met	Pro	Asn	Ile	Glu 535	Arg	Phe	Pro	Asn	Gly 540	Lys	Lys	Tyr	Glu	1632
40	Leu 545	Asp	Lув	Tyr	Thr	Met 550	Phe .	His	Tyr	Leu	Arg 555	Ala	Gln	Glu	Phe	Glu 560	1680
	His	Gly	Lys	Ser	Arg 565	Ile .	Ala :	Leu	Thr	<b>Asn</b> 570	Ser	Val	Asn	Glu	Ala 5 <u>7</u> 5	Leu	1728 1776
45	Leu .	Asn	Pro	Ser 580	Arg	Val '	Tyr '	Thr	Phe 585	Phe	Ser	Ser	Asp	Tyr 590	Val	Lys	
50	Lys '	Val	Asn 595	Lys	Ala	Thr (	Glu /	Ala . 500	Ala	Met	Phe	Leu	Gly 605	Trp	Val	Glu	1824
		610	Val	Tyr	Asp	Phe :	Thr 1 615	Asp (	3lu '	Thr	Ser (	Glu 620	Val	Ser	Thr	Thr	1872
55	GAT Asp 1	Lys	ATT Ile	GCG Ala	qaA	ATA A Ile 7 630	thr I	lle :	lle :	Ile :	CCA ' Pro ' 635	TAT I	ATA Ile	GGA Gly	Pro	GCT Ala 640	1920

	TT Le	A AA u As	T AT	A GG e Gl	T AA: y Asi 64!	a Me	G TTA	TA1	Lys	GAT S Asp 650	) Asp	r TTT	GTA Val	GG Gl	GC Y Ala 65	TTA Leu 5	196
5	AT.	A TT e Ph	T TC e Se	A GG r Gly 660	/ Ala	r GT: a Val	r ATT	CTC Leu	TTA Leu 665	ı Glu	TTI Phe	T ATA	CCA Pro	GAC Glu	ı Ile	GCA Ala	201
10	AT.	A CC	T GT/ o Va: 67!	l Lev	A GG1 u Gly	Thi	TTT Phe	GCA Ala 680	Leu	GTA Val	TCA Ser	TAT	ATT Ile 685	Ala	AA1 Asr	AAG Lys	2064
	GT. Val	CT. Lev 69	u Thi	C GTT	CAA Gln	ACA Thr	ATA Ile 695	Asp	AAT Asn	GCT Ala	TTA Leu	AGT Ser 700	Lys	AGA	AAT Asn	GAA Glu	2112
15	AAA Lys 705	TI	G GAT	GAG Glu	GTC Val	TAT Tyr 710	Lys	TAT Tyr	ATA Ile	GTA Val	ACA Thr 715	AAT Asn	TGG Trp	TTA Leu	GCA Ala	AAG Lys 720	2160
20	GTT Val	AA1 Asr	ACA Thr	CAG Gln	ATT Ile 725	GAT Asp	CTA Leu	ATA Ile	AGA Arg	AAA Lys 730	AAA Lys	ATG Met	AAA Lys	GAA Glu	GCT Ala 735	TTA Leu	2208
	GAA Glu	AAT Asn	CAA Gln	GCA Ala 740	GAA Glu	GCA Ala	ACA Thr	AAG Lys	GCT Ala 745	ATA Ile	ATA Ile	AAC Asn	TAT Tyr	CAG Gln 750	TAT Tyr	AAT Asn	2256
25	CAA Gln	TAT	Thr 755	Glu	GAA Glu	GAG Glu	AAA Lys	AAT Asn 760	AAT Asn	ATT Ile	AAT Asn	TTT Phe	AAT Asn 765	ATT Ile	GAT Asp	GAT Asp	2304
30	TTA Leu	AGT Ser 770	Ser	AAA Lys	CTT Leu	AAT Asn	GAG Glu 775	TCT Ser	ATA Ile	AAT Asn	AAA Lys	GCT Ala 780	ATG Met	ATT Ile	AAT Asn	ATA Ile	2352
	AAT Asn 785	AAA Lys	TTT Phe	TTG Leu	AAT Asn	CAA Gln 790	TGC Cys	TCT Ser	GTT Val	TCA Ser	TAT Tyr 795	TTA Leu	ATG Met	AAT Asn	TCT Ser	ATG Met 800	2400
35	ATC Ile	CCT Pro	TAT Tyr	GGT Gly	GTT Val 805	AAA Lys	CGG Arg	TTA Leu	Glu	GAT Asp 810	TTT Phe	GAT Asp	GCT Ala	AGT Ser	CTT Leu 815	AAA Lys	2448
	Asp	Ala	Leu	Leu 820	Lys	Tyr	ATA Ile	Tyr .	Asp 825	Asn .	Arg	Gly	Thr	Leu 830	Ile	Gly	2496
40	CAA Gln	GTA Val	GAT Asp 835	AGA Arg	TTA Leu	AAA Lys	GAT Asp	AAA Lys 840	GTT :	AAT . Asn .	AAT . Asn '	Thr	CTT : Leu : 845	AGT Ser	ACA Thr	GAT Asp	2544
45	ATA Ile	CCT Pro 850	TTT Phe	CAG Gln	CTT Leu	Ser	AAA ' Lys ' 855	TAC (	GTA ( Val :	GAT Asp	Asn (	CAA 1 Gln 1 860	AGA :	rta Leu	TTA Leu	TCT Ser	2592
				GAA Glu	Tyr								•				2613

50 (2) INFORMATION FOR SEQ ID NO: 8:

55

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 871 amino acids
- (B) TYPE: amino acid
  (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

5	Met Pro Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gl 1 5 10 15	
	Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pr 20 25 30	
10	Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Ar 35 40 45	3
	Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50	1
15	Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80	
	Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90 95	
20	Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val	
	Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys	
25	Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 135 140	
	Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 . 150 155 160	
30	Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175	
	Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190	
35	Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205	
	Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 220	
40	Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 230 235	
	Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255	
45	Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270	
	Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn 280 285	
50	Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300	
	Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 305 310 320	
55	Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 330 335	

	Ly	s Phe	Asp	Lys 340	Leu '	Tyr i	ys M	let I	eu I 45	hr G	lu I	le T	yr I	hr G 50	lu Asp
5	Ası	n Phe	Val 355	Lys	Phe I	Phe L	ys V 3	al L 60	eu A	sn Ai	rg Ly	/s Tl 36	nr T 55	yr L	eu Asn
	Phe	370	Lys	Ala	Val I	Phe L 3	ys I 75	le A	sn I	le Va	11 P1 36	TO L	/S Vi	al A	sn Tyr
10	Thr 385	Ile	Tyr	Asp (	Sly F	he A 90	sn L	eu A	rg A	sn Th 39	r As	n Le	u Al	a Al	a Asn 400
	Phe	Asn	Gly	Gln 4	Asn T	hr G	lu II	le As	sn As 41	n Me lo	t As	n Ph	e Th	r Ly 41	s Leu 5
15			•	-20				4.4	25				43	0	l Arg
			722				33	U				44	5		n Lys
20		Leu 450				4.5	2				460	)			
	403	Pro			* *	, ,				475	i				480
25		Thr		4	35				49	,				495	i
		Leu :	5	.00				509	5				510	1	
30			173				520	)				525			
	•	Leu M 530 Asp L				333	•				540				
35	545 His (				251	,				555					560
	Leu A			20	>				570					<b>5</b> 75	
40	Lys V		56					585					590		_
	Gln L	. 5	75				600					605			
45	Asp L	10			Ile	615 Thr				I	620				
	625 Leu A			y Asr	Met				Asp	635				,	640
50	Ile P		r G1	645 Y Ala			Leu	Leu	650				- 1	555	
	Ile P	ro Va 67	660 1 Le: 5		Thr	Phe		665 Leu	Val .	Ser 1			570 Ma <i>l</i>	Asn I	ys

	Va.	1 Le <sup>-</sup>	u Th:	r Val	l Glm	Thr	F Ile	Asp	Asr	Ala	a Lev	Ser 700	Lys	Arq	g Ası	a Glu
5	Lys 705	s Tr	p Ası	Glu	ı Val	710	Lys	туг	Ile	· Val	1 Thr 715	Asn	Trp	Let	ı Ala	Lys 720
	Val	ASI	Th:	Gl n	725	Asp	Leu	Ile	Arg	730	E Lys	Met	Lys	Glu	Ala 735	Leu
10	Glu	Asr	Gln	740	Glu	Ala	Thr	Lys	Ala 745	Ile	lle	Asn	Туг	Gln 750	Tyr	Asn
	Gln	Tyr	755	Glu	Glu	Glu	Lys	Asn 760	Asn	Ile	Asn	Phe	Asn 765	Ile	Asp	Asp
15	Leu	Ser 770	Ser	Lys	Leu	Asn	Glu 775	Ser	Ile	Asn	Lys	Ala 780	Met	Ile	Asn	Ile
	Asn 785	Lys	Phe	Leu	Asn	Gln 790	Cys	Ser	Val	Ser	Tyr 795	Leu	Met	Asn	Ser	Met 800
20	Ile	Pro	Tyr	Gly	Val 805	Lys	Arg	Leu	Glu	Asp 810	Phe	Asp	Ala	Ser	Leu 815	Lys
	Asp	Ala	Leu	Leu 820	Lys	Tyr	Ile	Tyr	Asp 825	neA	Arg	Gly	Thr	Leu 830	Ile	Gly
25			033					840			Asn		845			-
	Ile	Pro 850	Phe	Gln	Leu .	Ser	Lys 855	Tyr	Val .	Asp	Asn	Gln 2 860	Arg	Leu	Leu	Ser
80	Thr 865	Phe	Thr	Glu		11e : 870	Lys									
	(2) !	INFO	RMATI	ON FO	OR SE	Q ID I	VO: 9:	i								
5		(i) SE	QUEN	ICE C	HARA	CTER	ISTIC	S:								
		6	A) I EN	IGTH:	2628	hase	naire									

- 35
  - (A) LENGTH: 2628 base pairs(B) TYPE: nucleic acid

  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: DNA (genomic) (ix) FEATURE:

    - (A) NAME/KEY: CDS
    - (B) LOCATION:1..2628
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

55

40

45

	ATG Met 1	CAG Gln	TTC Phe	GTG Val	AAC Asn 5	AAG Lys	CAG Gln	TTC Phe	AAC Asn	TAT Tyr 10	AAG Lys	GAC Asp	CCT Pro	GTA Val	AAC Asn 15	GGT Gly	48
5	GTT Val	GAC Asp	ATT Ile	GCC Ala 20	TAC Tyr	ATC Ile	AAA Lys	ATT Ile	CCA Pro 25	AAC Asn	GCC Ala	G] y	CAG Gln	ATG Met 30	CAG Gln	CCG Pro	96
10	GTG Val	AAG Lys	GCT Ala 35	TTC Phe	AAG Lys	ATT Ile	CAT His	AAC Asn 40	AAA Lys	ATC Ile	TGG Trp	GTT Val	ATT Ile 45	CCG Pro	GAA Glu	CGC Arg	144
	GAT Asp	ACA Thr 50	TTT Phe	ACG Thr	AAC Asn	CCG Pro	GAA Glu 55	GAA Glu	GGA Gly	GAC Asp	TTG Leu	AAC Asn 60	CCG Pro	CCG Pro	CCG Pro	GAA Glu	192
15																	

	GC/ Ala 65	a Lys	G CAC	G GTG	CCA Pro	GTT Val 70	Ser	TAC Ty	TAC Ty	GA1	TC/ Sei	Th	C TA	r CT	G AG	C ACA r Thr 80	240
5	GA( Asp	AAC Asr	GAC Glu	AAG Lys	GAT Asp 85	Asn	TAC	CTC Lev	AAC Lys	G GGF Gly 90	' Val	ACC Thi	C AA/	A TT	9 TT	CGAG ∈ Glu	288
10					Thr					Met					: Ile	GTC Val	336
				Pro					Ser					Glu		AAG Lys	384
15			Asp										Asp			TAC	432
20	Arg 145	Ser	Glu	GAA Glu	Leu	Asn 150	Leu	Val	Ile	Ile	Gly 155	Pro	Ser	Ala	Asp	Ile 160	480
	Ile	Gln	Phe	GAG Glu	Cys 165	Lys	Ser	Phe	Gly	His 170	Glu	Val	Leu	Asn	Leu 175	Thr	528
25	Arg	Asn	Gly	TAC Tyr 180	Gly	Ser	Thr	Gln	Tyr 185	Ile	Arg	Phe	Ser	Pro 190	Asp	Phe	576
30				TTC													624
30	Gly	GCA Ala 210	Gly	Lys	TTC Phe	Ala	ACT Thr 215	GAT Asp	CCA Pro	GCG Ala	GTG Val	ACC Thr 220	CTG Leu	GCA Ala	CAC His	GAG Glu	672
35	Leu 225	Ile	His	GCC Ala	Gly	His 230	Arg	Leu	Tyr	Gly	Ile 235	Ala	Ile	Asn	Pro	Asn 240	720
	Arg	Val	Phe		Val 245	Asn	Thr	Asn	Ala	Tyr 250	Tyr	Glu	Met	Ser	Gly 255	Leu	768
40	Glu	Val	Ser	TTC Phe 260	Glu	Glu	Leu	Arg	Thr 265	Phe	Gly	Gly	His	Asp 270	Ala	Lys	816
45				AGC Ser			Glu										864
	Lys			GAT Asp		Ala :					Lys						912
50	Gly 305	Thr	Thr		Ser :	Leu ( 310	3ln	Tyr	Met	Lys .	Asn 315	Val	Phe	Lys	Glu	Lys 320	960
5 <i>5</i>	TAT Tyr	CTC Leu	CTA Leu	Ser (	GAA ( Glu ) 325	Asp ?	ACA '	TCT Ser	Gly	AAA Lys 330	TTT Phe	TCG Ser	GTA Val	Asp	AAA Lys 335	TTA Leu	1008

	AA: Ly:	A TTT s Phe	GAT Asp	Lys 340	. Leu	TAC	Lys	ATC Met	TT) Let 349	Th:	A GAG	G AT	T TA	C AC r Th 35	r Gl	G GAT u Asp	1056
5	AA' Asi	r TT1 n Phe	GT1 Val 355	. Lys	TTT Phe	TTT Phe	AAA Lys	Val	. Le	AA( 1 Asi	AGA Arg	A AAI g Ly:	A ACI	Ty	r TT	TAA E neA u	1104
10	TT	GAT Asp 370	Lys	GCC Ala	GTA Val	TTT Phe	AAG Lys 375	Ile	AA7 Asr	TATA	GT#	CCT Pro 380	Lys	GT/ Val	A AA: L Asi	TAC Tyr	1152
	Thr 385	Ile	Tyr	Asp	Gly	Phe 390	Asn	Leu	Arg	Asn	395	Asr	Leu	Ala	a Ala	AAC Asn 400	1200
15	Phe	. Asn	Gly	Gln	405	Thr	Glu	Ile	Asn	410	Met	Asn	Phe	The	Lys 415		1248
20	Lys	Asn	Phe	Thr 420	Gly	Leu	Phe	Glu	Phe 425	Tyr	Lys	Leu	Leu	Cys 430	Val	_	1296
	Gly	Ile	11e 435	Thr	TCT Ser	Lys	Thr	Lys 440	Ser	Leu	Asp	Lys	Gly 445	Tyr	Asn	Lys	1344
25	Ser	Ala 450	Asp	Gly	GCA Ala	Leu	Asn 455	Asp	Leu	Cys	Ile	Lys 460	Val	Asn	Asn	Trp	1392
30	Asp 465	Leu	Phe	Phe	AGT	Pro 470	Ser	Glu	Asp	Asn	Phe 475	Thr	Asn	ysb	Leu	Asn 480	1440
	Lys	Gly	Glu	Glu	ATT Ile 485	Thr	Ser	Asp	Thr	Asn 490	Ile	Glu	Ala	Ala	Glu 495	Glu	1488
35	Asn	Ile	Ser	Leu 500	GAT Asp	Leu	Ile	Gln	Gln 505	Tyr	Tyr	Leu	Thr	Phe 510	Asn	Phe	1536
40	Asp	Asn	Glu 515	Pro	GAA Glu	Asn	Ile	Ser 520	Ile	Glu	Asn	Leu	Ser 525	Ser	Asp	Ile	1584
40	Ile	Gly 530	Gln	Leu	GAA Glu	Leu	Met 535	Pro	Asn	Ile	Glu	Arg 540	Phe	Pro	Asn	Gly	1632
45	Lys 545	Lys	Tyr	Glu		Asp : 550	Lys '	Tyr	Thr	Met	Phe 555	His	Tyr	Leu	Arg	Ala 560	1680
	Gln	Glu	Phe	Glu	CAT (His (	Gly :	Lys :	Ser .	Arg	Ile 570	Ala	Leu	Thr	Asn	Ser 575	Val	1728
50	Asn	Glu	Ala	Leu 580	TTA L	Asn :	Pro :	Ser .	Arg 585	Val	Tyr	Thr	Phe	Phe 590	Ser	Ser	1776
55	Asp	Tyr	Val S95	AAG . Lys	AAA ( Lys \	Val 1	Asn l	Lys :	Ala	Thr	Glu	Ala.	GCT Ala 605	Met	Phe	Leu	1824

	GG Gl	C TG y Tr 61	p Va.	A GAI l Glu	A CAA	TTA Leu	GTA Val 615	Туз	GA' Asj	T TT P Ph	T AC e Th	C GA I As 62	p Gl	A AC u Th	T AG	C GAA r Glu	1872
5	GT: Va. 62!	i Se	r AC	r ACC	GAT Asp	AAA Lys 630	ATT Ile	GCC Ala	GA:	r AT	A AC e Th: 63!	r Il	A AT	T AT e Il	T CC e Pr	A TAT O Tyr 640	1920
10	AT/	A GG	A CCT	GCT Ala	TTA Leu 645	AAT Asn	ATA Ile	GG1 Gly	AAT Ast	TATO Met 650	Let	A TAT	r AA	A GA S As	T GA P As 65	TTTT P Phe	1968
	GTA Val	GG1	GCT Ala	TTA Leu 660	Ile	TTT Phe	TCA Ser	GGA Gly	GC1 Ala 665	(Va)	T ATT	CTC Lev	TT/	GAI G1: 670	ı Pho	T ATA	2016
15	CCA Pro	GAC Glu	Ile 675	Ala	ATA Ile	CCT Pro	GTA Val	TTA Leu 680	Gly	ACT Thr	TTT Phe	GCA Ala	CTI Leu 685	[Va]	TCJ Sez	TAT	2064
20	ATT Ile	GCG Ala 690	Asn	AAG Lys	GTT Val	CTA Leu	ACC Thr 695	GTT Val	CAA Gln	ACA Thr	ATA Ile	GAT Asp 700	Asn	GCT Ala	TTA Leu	AGT Ser	2112
	AAA Lys 705	AGA Arg	AAT Asn	GAA Glu	AAA Lys	TGG Trp 710	GAT Asp	GAG Glu	GTC Val	TAT Tyr	AAA Lys 715	Tyr	ATA Ile	GTA Val	ACA Thr	AAT Asn 720	2160
25	Trp	Leu	Ala	Lys	GTT Val 725	Asn	Thr	Gln	Ile	<b>Asp</b> 730	Leu	Ile	Arg	Lys	Lys 735	Met	2208
	A <b>AA</b> Lys	GAA Glu	GCT Ala	TTA Leu 740	GAA Glu	AAT ( Asn (	CAA Gln	GCA Ala	GAA Glu 745	GCA Ala	ACA Thr	AAG Lys	GCT Ala	ATA Ile 750	ATA Ile	AAC Asn	2256
30	Tyr	Gln	Tyr 755	Asn	CAA Gln	Tyr :	Chr (	Glu 760	Glu	Glu	Lys	Asn	Asn 765	Ile	Asn	Phe	2304
35	AAT Asn	ATT Ile 770	gat Asp	gat Asp	TTA . Leu !	Ser S	rcg i Ser i	AAA Lys	CTT Leu	TAA Asn	GAG Glu	TCT Ser 780	ATA Ile	AAT Asn	AAA Lys	GCT Ala	2352
	ATG Met 785	ATT Ile	AAT Asn	ATA . Ile	AAT A	AAA 7 Lys E 790	TT 1 he I	TG . Leu .	AAT Asn	Gln	TGC Cys 795	TCT Ser	GTT Val	TCA Ser	TAT Tyr	TTA Leu 800	2400
40	ATG Met	TAA neA	TCT Ser	Met	ATC ( Ile 1 805	CT I	AT G	GT (	Val	AAA Lys 810	CGG Arg	TTA Leu	GAA Glu	Asp	TTT Phe 815	GAT Asp	2448
45	GCT Ala	AGT Ser	Leu	AAA ( Lys 2 820	GAT (	CA T	TA I eu L	eu 1	AAG Lys 325	TAT . Tyr	ATA Ile	TAT Tyr	Asp .	AAT Asn 830	AGA Arg	GGA Gly	2496
	ACT Thr	Leu	ATT ( Ile ( 835	GGT (	CAA G 31n V	TA G	sp A	GA 7	CTA :	AAA ( Lys )	GAT : Asp :	Lys '	GTT : Val : 845	AAT . Asn .	AAT Asn	ACA Thr	25 <b>44</b>
50	CTT :	AGT . Ser '	ACA ( Thr i	GAT A	ATA C	TO P	TT C he G 55	AG (	eu !	rcc : Ser 1	Lys :	TAC ( Tyr 1 860	JTA ( Val )	Asp A	AAT ( Asn (	CAA 31n	2592
55	AGA 1 Arg 1 865	TTA : Leu :	ITA : Leu :	rcr A Ser T	hr P	TT A he T 70	CT G. hx G	AA I lu I	yr I	lle I	AAG : Lys 975	TAA *					2628

	(i) SEQUENCE CHARACTERISTICS:
5	<ul><li>(A) LENGTH: 876 amino acids</li><li>(B) TYPE: amino acid</li><li>(D) TOPOLOGY: linear</li></ul>
40	(ii) MOLECULE TYPE: protein
10	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:
15	
20	
25	
30	
35	
40	
45	
50	

55

(2) INFORMATION FOR SEQ ID NO: 10:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro 20 25 30 Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg 35 40 Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 10 Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 95 95 15 Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys 20 Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 140 Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 25 Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 180 30 Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205 Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 220 35 Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 230 235 Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255 40 Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270 Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn 275 280 285 45 Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 300

50

		Gl <sub>3</sub> 305		Thi	c Ala	a Se	7 Let 310		ту:	r Met	Ly:	315		l Phe	E Ly	s Gl	u Lys 320
5		Туг	: Lev	Let	ı Ser	329		The	: Sei	Gly	/ Lys		Se1	: Val	Asp	331	s Leu 5
		Lys	Phe	Asp	340		туг	Lys	Met	1 Leu 345		Glu	Ile	ту:	7h1		u Asp
10		Asn	Phe	Val 355		Phe	Phe	Lys	360		. Asn	Arg	Lys	Th: 365		Let	ı Asn
		Phe	Asp 370		Ala	Val	Phe	Lys 375		Asn	Ile	Val	Pro 380		Val	Asr	Tyr
15		Thr 385		Tyr	Asp	Gly	Phe 390		Leu	Arg	Asn	Thr 395	Asn	Leu	Ala	Ala	Asn 400
		Phe	Asn	Gly	Gln	Asn 405		Glu	Ile	Asn	Asn 410	Met	Asn	Phe	Thr	Lys 415	Leu
20		Lys	Asn	Phe	Thr 420	Gly	Leu	Phe	Glu	Phe 425	Tyr	Lys	Leu	Leu	Сув 430	Val	Arg
		Gly	Ile	Ile 435	Thr	Ser	Lys	Thr	Lys 440	Ser	Leu	Asp	Lys	Gly 445	Tyr	Asn	Lys
25		Ser	Ala 450	Asp	Gly	Ala	Leu	Asn 455	Asp	Leu	Cys	Ile	Lys 460	Val	Asn	Asn	Trp
		465					Pro 470			_		475			-		480
30		-	•			485	Thr	_			490					495	
					500	-	Leu			505	-	-			510		
35				515			Asn		520					525		-	
			530				Leu	535					54Ŏ				•
40	:	545	-	•			Asp 550					555		-		_	560
40	(	Gln	Glu	Phe		His 565	Gly	Lys	Ser		11e 570	Ala	Leu	Thr		Ser 575	Val
	i	Asn	Glu		<b>Le</b> u 580	Leu	Asn	Pro		Arg 585	Val	Tyr	Thr		Phe 590	Ser	Ser
45	,	Asp '	-	Val 595	Lys	Lys	Val .		Lys	Ala	Thr	Glu i		Ala 605	Met	Phe	Leu
	(		Trp 610	Val (	Glu	Gln	Leu	Val 615	Tyr	Asp	Phe '		Asp (	Glu	Thr :	Ser	Glu
50		Val :	Ser '	Thr '	Thr .	Asp	Lys 630	Ile .	Ala .	Asp :		Thr :	Ile :	Ile	Ile :		Tyr 640
	3	(le (	Gly :	Pro 2		Leu 645	Asn :	Ile (	Gly .		Met 1 650	Leu 7	(yr	Lys i		4 <b>s</b> p 555	Phe

	Va.	1 G1;	y Ala	660	ı Ile	e Pho	e Sei	Gl;	y Al. 66.	a Va S	1 11	e Le	u Lei	u Gl 67		e Il
5	Pro	Gl:	1 Ile 675	Ala	Ile	Pro	Va]	Le:	u Gl	y Th	r Phe	= Ala	Let 685		l Se	r Tyı
	Ile	Ala 690	Asn	Lys	Val	Let	1 Thr 695	· Val	Glr	Th:	r Ile	700		Ala	a Let	ı Ser
10	705	•				/10					715					720
					/25		Thr			730	,				735	
15				740			Gln		745					750		
	Tyr	Gln	Tyr 755	Asn	Gln	Tyr	Thr	Glu 760	Glu	Glu	Lys	Asn	Asn 765	Ile	Asn	Phe
20	Asn	11e 770	Asp	Asp	Leu	Ser	Ser 775	Lys	Leu	Asn	Glu	Ser 780	Ile	Asn	Lys	Ala
	Met 785	Ile	Asn	Ile	Asn	Lys 790	Phe	Leu	Asn	Gln	Cys 795	Ser	Val	Ser	Tyr	Leu 800
25	Met	Asn	Ser	Met	Ile 805	Pro	Tyr	Gly	Val	Lys 810	Arg	Leu	Glu	Asp	Phe 815	Asp
	Ala	Ser	Leu	Lys . 820	Asp	Ala	Leu	Leu	Lys 825	Tyr	Ile	Tyr	Asp	Asn 830	Arg	Gly
30			835					840					845			
	Leu	Ser 850	Thr i	Asp :	Ile	Pro	Phe 855	Gln	Leu	Ser	Lya	<i>Tyr</i> 860	Val .	Asp	Asn	Gln
35	Arg :	Leu :	Leu S	Ser 7		Phe 970	Thr	Glu '	Tyr		Lys 875	•				

- (2) INFORMATION FOR SEQ ID NO: 11:
- 40 (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 2637 base pairs

  - (B) TYPE: nucleic acid
    (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
  - (ii) MOLECULE\_TYPE: DNA (genomic)
  - (ix) FEATURE:

45

50

55

- (A) NAME/KEY: CDS
- (B) LOCATION:1..2637
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

	ATG Met 1	CAG Gln	TTC Phe	GTG Val	AAC Asn 5	AAG Lys	CAG Gln	TTC Phe	AAC Asn	TAT Tyr 10	AAG Lys	GAC Asp	CCT Pro	GTA Val	AAC Asn 15	GGT Gly	48
5	GTT Val	GAC Asp	ATT Ile	GCC Ala 20	TAC Tyr	ATC Ile	AAA Lys	Ile	CCA Pro 25	AAC Asn	GCC Ala	GGC Gly	CAG Gln	ATG Met 30	CAG Gln	CCG Pro	96
10																	
15																	
20																	
25																	
30														-			
35																	
40																	
45																	
50																	

	G: Va	rg A	AG ( ys )	SCT ( Mail 35	TTC I	AAG . Lys :	ATT (	IIS A	AC A Sn L 40	AA I	ATC Lle	TGG Trp	GTT Val	AT	Pr	CG G	AA lu	CGC Arg	144
5	GA As	p I	CA T hr F 50	TT A	ACG A	AAC ( Asn I	ro G	AA G lu G 55	AA G lu G	GA (	SAC Asp	TTG Leu	AAC Asn 60	CCC	G CC	G C	CG	GAA Glu	192
10	6	а Ly 5	ys G	ID A	al F	ro v	TT T al S 70	er 1	yr T	yr A	sp :	Ser 75	Thr	Tyr	Le	u S	er '	Thr 80	240
	A.S	p As	sn G	ıa r	ys A	85 85	AC T	AI PO	eu Ly	ys G	1y \ 90	/al	Thr	Lys	Le	u Ph	ne (	31u	288
15	Arg	3 11	.e 17	YF S	er T	nr A	AC CT	an G1	10	:g M: 15	et I	eu :	Leu	Thr	Se:	r Il	e v	<b>Tal</b>	336
20	Arg	i GI	y 1.	is P	ro P	ne T	GG GG	y G1 12	y Se	r T	ar I	le ;	<b>Asp</b>	Thr 125	Glu	. Le	u L	ys	384
	val	13	<b>e</b> As	ip Ti	T A	sn C	C AT	e As 5	n Va	1 11	le G	ln I	40	Asp	Gly	Se	T	yr	432
25	145	s se:	r GI	u G1	.u Le	15		u Va	1 11	e Il	e G.	ly P	ro s	Ser	Ala	Asy	10	le 60	480
30	116	GII	ı Pn	e GI	16	s Ly	G AG s Se:	r Phe	e Gl	y Hi 17	s G] 0	lu V	al I	.eu	Asn	Let 175	T?	ır	528
	Arg	ASI	f GT	y Ty 18	O F GI	y se	Th:	r Glī	185	r Il	e Ar	g P	he S	er	Pro 190	Asp	Ph	ie.	576
35	Thr	Pne	19	y PR	e GI	u GI	ACI	200	Glu	Va.	l As	p Ti	ar A 2	en 1 05	Pro	Leu	Le	u ·	624
40	GIÀ	210	GT?	Lyı	Phe	e Ala	Thr 215	Asp	Pro	Ala	ı Va.	1 Th 22	r L	eu A	lla	His	Gl	u	672
	Leu 225	Ile	His	Ala	r GT	230	Arg	Leu	Tyr	Gly	23!	e Al 5	a I	le A	sn	Pro	24 (	n D	720
45	Arg	Val	Phe	Lys	245	. Ast	Thr	Asn	Ala	7yr 250	Туз	r Gl	u Me	et S	er (	31y 255	Let	1	768
50	GAA (	Val	ser	260	Glu	Glu	Leu	Arg	Thr 265	Phe	Gly	r Gl	y Hi	.s A	8p 7	lla	Lys	•	816
50	Phe	Ile	275	Ser	Leu	Gln	Glu	Asn 280	Glu	Phe	Arg	Let	28	T T)	/ <b>=</b> 1	yr	Asn	•	864
55	Lys I	he 290	AAA Lys	gat Asp	Ile	Ala	AGT Ser 295	ACA Thr	CTG Leu	AAC Asn	AAG Lys	Ala 300	Ly:	G TO	C A	TT le	GTG Val		912

	GG Gl 30	y m	CC AC	ET GO	CT TO	A TT r Le 31	u GI	G TA n Ty	T Al	G A	/S A	AT ( sn \ 15	STT : /al	Phe	AAA Lys	GAC Glu	AAA Lys 320		960
5	ту	r Le	u Le	u Se	32	u Asy 5	p Th:	r Se:	r Gl	у L <sub>3</sub>	e P	he S	er V	al P	sp	Lys 335		1	800.
10	Lys	, Pu	e As	9 Ly 34	o O	A TAG u Ty:	r Lys	. Met	34	u Th 5	r G	lu I	le T	уг 1 3	hr ( 50	3lu	Asp	1	056
	Asn	ו אם	e va 35	I Ly	s Pn	r rra	Lys	360	l Lei	u As	n Ai	rg L	ys T	hr T 65	yr I	Leu	Asn	1	104
15	Pne	370	o P rA:	s Al	a va.	A TTI L Phe	375	Ile	: Ası	n Il	e Va	11 Pi 36	ro Ly	ys V	al A	sn	Tyr	1:	152
20	385	116	2 TY1	. ASI	5 G13	Phe 390	Asn	Leu	Arg	j Ası	1 Th	r As 5	n Le	eu Al	a A	la	Asn 400	12	200
	Pne	AST	ı Gış	GII	405		GLu	Ile	Asn	410	i Me	t As	n Ph	e Th	r L	ys 15	Leu	12	48
25	Lys	ASN	Phe	420	Gly	TTG	Phe	Glu	Phe 425	Туг	Ly	s Le	u Le	u Cy 43	s Va	al.	Arg	12	96
30	GIA	TIE	435	Thr	Ser	AAA Lys	Thr	Lys 440	Ser	Leu	Ası	) Ly	44	у Ту 5	r As	in i	Lys	13	44
	116	450	GIA	Arg	Cys	GAT Asp	Gly 455	Ala	Leu	Asn	Asp	46(	Cy:	s Il	e Ly	s \	/al	139	92
35	AAT Asn 465	ASN	Trp	Asp	Leu	Phe 470	Phe	Ser	Pro	Ser	Glu 475	Asp	) Ası	n Pho	Th	I A	sn 80	144	10
40	GAT Asp	Leu	Asn	Lys	485	Glu	Glu	Ile	Thr	Ser 490	Asp	Thr	Asr	Ile	G1 49	u A 5	la	148	8
	GCA (	31 II	Glu	500	Ile	Ser :	Leu i	Asp	Leu 505	Ile	Gln	Gln	Tyr	Tyr 510	Lei	u T	hr	153	6
45	Phe	Asn	Phe 515	Asp	Asn	Glu 1	Pro (	31u 2 520	Asn	Ile	Ser	Ile	Glu 525	Asn	Le	ı S	er	158	4
50		sp 30	Ile	Ile	Gly (	Gln i	35	ilu I	Leu i	Met	Pro	Asn 540	Ile	Glu	Arg	Pi	ne	1632	2
	CCT A Pro A 545	sn (	Gly :	Lys	Lys :	Tyr 0	Slu I	eu A	\sp	Lys	Tyr 5 <b>55</b>	Thr	Met	Phe	His	7) 56	)T 50	1680	)
55	CTT C Leu A	rg i	GCT ( Ala (	Gln (	GAA : Glu I 565	Phe G	AA C	AT G	HY I	AAA Lys : 570	TCT Ser	AGG Arg	ATT Ile	GCT Ala	TTA Leu 575	Th	IA Ir	1728	t

	A) A:	AT T	CT ( Ser \	al P	AAC ( Asn ( 80	GAA ( Glu /	GCA T Ala L	TA T eu L	eu A	AT C sn P 85	CT A	GT C	rg V	al T	TAT A	CA hr	TTT Phe	1776
5	Pf	ie S	er S	er A 95	sp 1	yr v	ITA A /al L	ys L	ys V	al A	sn L	ys A	la T	hr G 05	lu A	1a .	Ala	1824
10	MG	6	ne 1	eu G	TA I	rp v		14 G. 15	in Le	eu V	al T	yr A:	sp Pl 20	ne T	hr A	ga (	Glu	1872
45	7n 62	r S0 5	er G	lu V	al S	er T	CT AC hr Tl 30	ur As	sp L	/S I	le A: 6:	la As 35	sp Il	le Ti	ar I	le 1	[le 540	1920
15	110	e Pi	ro T	/T 1.	Le G.	1y P:	CT G( ro A)	la Le	eu As	in I) 65	le G:	ly As	n Me	t Le	65 65	/r L	'À a	1968
20	Ası	) AS	ib fi	1e Va	60 60	LY A.	TT TT	u II	e Ph 66	e Se 5	r Gl	y Al	a Va	1 I1 67	e Le	u L	eu	2016
05	GLU	ı Pn	e II 67	6 Pr	o Gi	u II	T GC	a II 68	e Pr O	o Va	l Le	u Gl	y Th 68	r Ph S	e Al	a L	eu	2064
25	Val	69	r 1y	Y 11	e Al	a As	T AA D Ly 69	s Va. S	l Le	u Th	r Va	1 Gl:	n Th: 0	r Il	e As	p A	8n	2112
30	705 GTA	AC	A AA	r Ly r TG	s ar G TT	g As 71 A GC	n Gli O A AAG	u Ly:	s Tr	r aci	71:	u Vai	l Tyi	· Ly:	з Ту:	72	le 20	2160
	VA1 AAA	AA	ATY	a Tir	P Le 72: A GA	A GC	a Ly: r rr:	S Val	L Asr	730	r Gli	l Ile	. Asp	Let	739	AI	rg m	2208
35	Lys ATA	LYS	AAC	740 TA	G CAC	I AI	a Let Caat	CAA	ASE 745	Glr ACT	Ala GAG	Glu	Ala GAG	The 750	Lys	Al	a T	2304
40	ATT	AAT	755 TTI	AA1	. Gir	Ty:	: Asn	Gln 760	AGT	Thr	Glu	Glu	765	Lys	Asn	As	n N	2352
	AAT	AAA	GCT	ATG	ATI	TAA	Asp 775	AAT	AAA	TTT	TTG	780	CAA	TCC	Jr Contr	Collect	<b>.</b>	2400
45	785 TCA	TAT	TTA	ATG	AAT	790 TCT	ATG	ATC	CCT	ТАТ	795 GGT	ململت		ccc	TYPA	800		2448
50	GAT	TYF TTT	GAT	GCT	805	CTT	AAA	GAT	Pro GCA	BIO TTA	Gly	Val	Lys	Arg	Leu 815	Glu		2496
	ASP	Pne AGA	ASP GGA	820 ACT	SET	Leu	Lys	CAA	Ala 825 GTA	Leu	Leu	Lys	Tyr	Ile 830	Tyr	Asp	•	2544
55	Asn	Arg	Gly 835	Thr	Leu	Ile	Gly	Gln 840	Val	Asp	Arg	Leu	Lys 845	Asp	Lys	Val		-21

5	GAT	850	CAA	CTT Leu AGA Arg	TTA	TTA	855	11e	ALO.	Pne	GAA Glu	Leu 860	Ser	Lys	Tyr	GTA Val	2592 2637
10	(2) ا			ION FO		≣Q ID					875						
15		( <i>A</i>	4) LEN 3) TYF	NCE C NGTH: PE: am POLO	: 879 a	amino cid											
20				NCE [				EQ IC	NO:	12:							
25																	
30																	
35																	
40																	
45																	
50																	

		_														
		_			-	•				1	0				1	-
5	Va.	l As	p Ile	e Ala 20	Tyz	: Ile	. Lys	Ile	2 Pro	Ası 5	n Ala	a Gly	/ Gl	n Met		n Pro
	Va.	l Ly	a Ala 35	Phe	Lys	Ile	His	Asr 40	Lys	s Ile	: Trj	Val	1 Ile		Gl	ı Arg
10	Ası	Th: 50	Phe	Thr	Asn	Pro	Glu 55	Glu	gly	/ Asr	Lev	Asn 60		Pro	Pro	Glu
	Ala 65	Lys	Glr	Val	Pro	Val	Ser	Тух	Тут	. yat	Ser 75	Thr	Туг	Leu	Sex	Thr 80
15	Asp	) Asr	Glu	Lys	Asp 85	Asn	Tyr	Leu	Lys	Gly 90	Val	Thr	Lys	Leu	Phe 95	
	Arg	Ile	Tyr	Ser 100	Thr	Asp	Leu	Gly	Arg 105	Met	Leu	Leu	Thr	Ser 110	Ile	Val
20	Arg	Gly	Ile 115	Pro	Phe	Trp	Gly	Gly 120	Ser	Thr	Ile	Asp	Thr 125	Glu	Leu	Lys
	Val	Ile 130	Asp	Thr	Asn	Cys	Ile 135	Asn	Val	Ile	Gln	Pro 140	Asp	Gly	Ser	Tyr
25	Arg 145	Ser	Glu	Glu	Leu	Asn 150	Leu	Val	Ile	Ile	Gly 155	Pro	Ser	Ala	Asp	Ile 160
	Ile	Gln	Phe	Glu	Сув 165	Lys	Ser	Phe	Gly	His 170	Glu	Val	Leu	Asn	Leu 175	Thr
30	Arg	Asn	Gly	Tyr 180	Gly	Ser	Thr	Gln	Tyr 185	Ile	Arg	Phe	Ser	Pro 190	Asp	Phe
	Thr	Phe	Gly 195	Phe ·	Glu	Glu	Ser	Leu 200	Glu	Val	Asp	Thr	Asn 205	Pro	Leu	Leu
35	Gly	Ala 210	Gly	Lys	Phe .	Ala	Thr . 215	qaA	Pro	Ala	Val	Thr 220	Leu	Ala	His	Glu
	Leu 225	Ile	His	Ala (	Gly	His 230	Arg :	Leu	Tyr		Ile 235	Ala	Ile	Asn		Asn 240
40	Arg	Val	Phe	Lys \	Val 2	Asn '	Thr I	Asn .		Tyr 250	Tyr	Glu i	Met			

	Glu	Val :	Ser E	he Gl	u Gli	u Leu	Arg	Thr 265	Phe (	Gly G	ly Hi	3 Asp 270	Ala Lys
5	Phe	Ile A	Asp S 275	er Le	u Glr	n Glu	Asn 280	Glu	Phe A	arg Le	u Ty:	Tyr	Tyr Asn
	Lys	Phe I 290	ys A	sp Il	e Ala	Ser 295	Thr	Leu	Asn I	ys Al 30		Ser .	Ile Val
10	Gly 305	Thr T	hr A	la Se:	7 L <b>e</b> u 310	Gln	Tyr	Met	Lys A 3	sn Va 15	l Phe	Lys (	3lu Lys 320
				34:	<b>ο</b> ,				330			3	ys Leu 35
15			34	10				345	•			350	lu Asp
		3:	22				360				365		eu Asn
20		370				3/5				380	•		sn Tyr
25	363				390				35	5			la Asn 400 ys Leu
25				405				4	10			4:	ys Leu 15 11 Arg
<i>30</i>		le Il	42 e Th:	U			4	25				430 Tyr As	_
	Ile G	43	J		Asp (	Gly A	140			p Ľeu	445	Ile Ly	
35			p Asp	Leu		45 <b>5</b> Phe S	Ger P	ro Se	r Gl:		Asn I	Phe Th	
	Asp L	eu Ası	ı Lys	Gly 485	Glu (	3lu I	le T	hr Se 49	r Asp		Asn 1	le Gl 49	
40	Ala G	lu Glu	Asn 500	Ile	Ser I	Leu A	ap Le	eu Il 95	e Glr	Gln			
	Phe As	sn Phe 515	Asp	Asn (	Glu P	To G	lu As 20	n Il	e Ser	Ile	Glu A 525	sn Lei	ı Ser
45	Ser As	p Ile	Ile	Gly	Sln L 5	eu G 35	lu Le	u Me	t Pro	Asn 540	Ile G	lu Arg	Phe Phe
	Pro As 545			=	350				555				560
50	Leu Ar			202				570	,			575	
	Asn Se		580				58	5			5.5	90	
55	Phe Se	r Ser 595	Asp	Tyr V	aı Ly	60 Ey	s Val	l Asn	Lys		hr Gl	u Ala	Ala

	М	et P	he L. 10	eu G	ly Tr	p Va	l Gl: 615	ı G1.	n Le	u Va	1 ту	r As; 62:	p Ph O	e Th	ır As	p Glu
5							_				635	•				e Ile 640
						_				031	,				65	-
10					-				003	,				670	0	ı Leu
	G1	u Ph	e Il 67	e Pr	o Glu	ı Ile	Ala	Ile 680	Pro	Val	. Leu	Gly	Th: 685	Phe	e Ala	Leu
15	Va	1 Se 69	r Ty:	r Ile	e Ala	Asn	Lys 695	Val	Leu	Thr	Val	Gln 700	Thr	Ile	Asp	Asn
	Ala 70:	a Le:	u Se:	r Lys	a Arg	Asn 710	Glu	Lys	Trp	Asp	Glu 715	Val	Tyr	Lys	Tyr	Ile 720
20	Va]	l Thi	r Asr	ı Trp	1 Leu 725	Ala	Lys	Val	Asn	Thr 730	Gln	Ile	Asp	Leu	Ile 735	Arg
									/45		Ala	•		750		
25	-							760			Glu		765			
							,,,					780				
30											Leu / 795					800
					003				•	RTO	Gly (				815	
35								•	,25		Leu L			830		
							٥	40			Arg L	8	45			
40						٠	33					60			yr v	'al
	Asp . 865	naA	Gln .	Arg 1	Leu I	Leu S 170	er T	hr P	he T	hr G	lu T <sub>1</sub> 75	Yr I	le L	λa	*	
45	(2) INFO	RMAT	ION F	OR SE	EQ ID i	NO: 13	₹•									

- - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 2862 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: double

    - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: DNA (genomic)
- (ix) FEATURE:

50

55

(A) NAME/KEY: CDS

## (B) LOCATION:1..2862

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

	A' Mo	TG C et G	AG (	rrc Phe	GTG Val	AAC Asn 5	AAC Lys	G CA B Gl	G T	TC A	AC 1	TAT Yr 10	AAG Lys	GA As	IC C	CT G	TA	AA( Asr 15	GGT Gly	•	48
5	G1 Va	MT G	AC A	ATT	GCC Ala 20	TAC Tyr	ATC Ile	AA Ly	A AT s Il	Le P	CA A ro A 25	AC sn	GCC Ala	GG G1	y G	ln M	TG et 30	CAG Gln	CCG Pro	•	96
10	va	IT E	ys A	35	Pne	Lys	IIe	Hi	s As 4	n Ly	/s I	le	Trp	۷a	1 11	le P:	ro (	Glu	CGC		144
	AS	p 11	nr P	ne :	rnr .	ASN	Pro	G10	2 7 GT	u G1	уА	sp 1	Leu	Ası 60	n Pr	O Pi	o I	Pro	GAA Glu		192
15	6 A1	а L} 5	/S G	ın v	/al :	Pro	70	Ser	ту	т Ту	r A	sp s	75	Thi	ту	r Le	u S	er	ACA Thr 80		240
20	AŞĮ	P AS	in G	iu L	ys /	85 85	ASN	тух	Let	u Ly	s Gi	y V	al	Thi	Ly	s Le	u P	he 95			288
	Arg	3 11	e Ty	/r s 1	er 1	ACT Thr	ASP	Leu	GI	10	g Me 5	t L	eu	Leu	Th:	r Se 11	r I O	le	Val		336
25	Arg	l GT	y 11	.e P: .5	ro F	he AC	Irp	GIÀ	Gly 120	Se:	r Th	r I	le i	Asp	Th:	Gl:	ı L	eu	Lys		384
30	AGA	130 TC:	e As O Ga	p TI	ar A	TT :	LYS	11e 135 CTC	ASN GTA	Val	. Il	e Gi	ln i	Pro 140	Asp	Gly	, Se	er '	Tyr		432
	145 ATC	CAC	r Gl	u Gi TGA	LU L	eu A 1 GC A	SO AG	Leu AGC	Val	GGC	CAC	e G1 15	Ly E	erc	Ser	Ala	As	p :	lle 160		528
35	CGT	AAC	GG:	e Gi	.u C;	ys L 65 3C T	ys :	ser	Phe	Gly	170	G G G	u V	al TC	Leu	Asn	Le 17	u 1	rhr		576
40	Arg	ASN	GG	, Ty 18	r G. O C GJ	ly s AG G	er 7 AG <i>1</i>	inr Agc	Gln CTG	Tyr 185 GAG	Ile	: Ar	g P T A	cc pe	Ser	Pro 190	As	p F	he TC		624
	GGT	Phe GCA	G13 195 GG0	Ph AA	e GI G TI	lu G	lu s Ca A	cr (	Leu 200 GAT	Glu	Val GCG	As	PT:	hr cc	Asn 205 CTG	Pro	Lei	u L	eu		672
45	CIG	A1a 210 ATC	CAC	GC	s Pn c GG	ie A. ST C	la T 2	inr i 15 GT (	Asp CTG	Pro	Ala	۷a ۵۳	1 T1 2:	hr 20	Leu arr	Ala	His	3 G	lu NG		720
50	Leu 225 CGC	Ile	TTC	Ala	a Gl	у н: 2: Т А	LS A 30 AC A	rg 1 cc 1	Leu AAC (	Tyr	Gly	116 23!	a Al	la :	Ile	ASD	Pro	2 A	sn 40 ra		768
50	GAA (	Val GTA	Phe AGC	Lys	24 24 : GA	l As 5 G GA	in T	hr A	Asn .	Ala ACG	Tyr 250 TTC	Tyr	. G)	iu i	Met TAT	Ser	G1y 255	r Le	eu ve		316
55	Glu 1	Val	Ser	Phe 260	: G1	u Gl	u L	eu A	irg :	Thr 265	Phe	Gly	ĞÌ	y i	lis .	Asp 270	Ala	L	rs	•	

	TTT ATC GAC AGC TTG CAG GAG AAC GAG TTC CGT CTG TAC TAC TAC AAC Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn 275 280 285	864
5	AAG TTT AAA GAT ATT GCA AGT ACA CTG AAC AAG GCT AAG TCC ATT GTG Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300	912
10	GGT ACC ACT GCT TCA TTA CAG TAT ATG AAA AAT GTT TTT AAA GAG AAA Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 305 310 320	960
	TAT CTC CTA TCT GAA GAT ACA TCT GGA AAA TTT TCG GTA GAT AAA TTA Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 330 JJS	1008
15	AAA TTT GAT AAG TTA TAC AAA ATG TTA ACA GAG ATT TAC ACA GAG GAT Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340 345 350	1056
20	AAT TTT GTT AAG TTT TTT AAA GTA CTT AAC AGA AAA ACA TAT TTG AAT Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355 360 3777 CAT AAA GTA TTT AAA GTA CTT AAC AGA AAA ACA TAT TTG AAT Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 365	1104
	TTT GAT AAA GCC GTA TTT AAG ATA AAT ATA GTA CCT AAG GTA AAT TAC Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 375 380	1152
25	ACA ATA TAT GAT GGA TTT AAT TTA AGA AAT ACA AAT TTA GCA GCA AAC Thr lle Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 395  TTT AAT GGT CAA AAT ACA GAA ATT AAT AAT ATG AAT TTT ACT AAA CTA	1200
30	AAA AAT TIT ACT GGA TTG TIT GAA TIT TAT ARG TITG CTD TOTAL AND TITG TITG TITG TITG TITG TITG TITG TIT	1248
	420 425 Leu Leu Cys Val Arg 420 425 GGG ATA ATA ACT TOT AAA ACT AAA TCA TTA GAT AAA CCT AAA ACT AAA ACT AAA TCA TTA GAT AAA CCT AAA ACT AAA ACT AAA TCA TTA GAT AAA CCT AAA ACT AAA ACT AAA TCA TTA GAT AAA CCT AAA ACT AAA ACT AAA TCA TTA GAT AAA ACT AAA ACT AAA ACT AAA TCA TTA GAT AAA ACT AAA ACT AAA ACT AAA ACT AAA ACT AAA TCA TTA GAT AAA ACT AAA ACT AAA ACT AAA ACT AAA ACT AAA TCA TTA GAT AAA ACT AAA AC	1296
35	ATC GAA GGT CGT TGC GAT GGG GCA TTA AAT GAT TTA TGT ANG AND GTT	1344
40	450 455 460	1392
40	AAT AAT TGG GAC TTG TTT TTT AGT CCT TCA GAA GAT AAT TTT ACT AAT Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 465 470 480	1440
45	GAT CTA AAT AAA GGA GAA GAA ATT ACA TCT GAT ACT AAT ATA GAA GCA Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 485 490 495	1488
	GCA GAA GAA AAT ATT AGT TTA GAT TTA ATA CAA CAA TAT TAT TTA ACC Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 505 510	1536
50	TTT AAT TIT GAT AAT GAA CCT GAA AAT ATT TCA ATA GAA AAT CTT TCA Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 515 520 525	1584
55	AGT GAC ATT ATA GGC CAA TTA GAA CTT ATG CCT AAT ATA GAA AGA TTT Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe 530 540	1632

5	CC Pr 54	o As	T GG n Gl	A AA y Ly	A AA s Ly	G TA s Ty 55	r Gl	G TT	A GA Li As	T AA. p Ly.	A TA' s Ty: 55:	r Th	T AT	G TI t Ph	C CA	TAT S Tyr	
·	CT'	r CG	T GC g Al	T CA a Gl	A GA n Gl: 56	u Pho	T GA	A CAT u His	r GG	T AA y Ly: 57	s Se	r Ar	G AT	r gc e Al	T TT a Le 57	A ACA u Thr	1728
10	AA: Asi	r TC	r GT r Va	T AA 1 As: 58	n Glu	A GCI Ala	A TT/	A TTA	AAA Asi S8:	n Pro	r AG:	r CG	T GT	Г ТА L Ту 59	r Th	A TTT r Phe	1776
15	TT1 Phe	TC: Sei	TC: Se: 59!	r Ası	TA1	GTA Val	A AAC Lys	AAA Lys 600	Va]	r AA7 L Asr	Lys	A GCT	T ACC Thi 605	. GI	G GC u Al	A GCT a Ala	1824
15	ATC Met	Phe 610	: Le	A GGG	TGC Trp	GTA Val	GAA Glu 615	ı Gln	TTA Leu	GTA Val	TAT Tyr	GAT Asp 620	) Phe	ACe Th	C GA	T GAA p Glu	1872
20	ACT Thr 625	Ser	GAJ Glu	A GTA	AGI Ser	Thr 630	Thr	GAT Asp	Lys	ATT Ile	GCG Ala 635	Asp	T ATA	AC:	r AT	A ATT E Ile 640	1920
	ATT Ile	Pro	TAT Tyr	' ATA	GGA Gly 645	Pro	GCT Ala	TTA Leu	AAT Asn	ATA Ile 650	Gly	AAT Asn	ATG Met	TTA	TA1	Lys	1968
25	GAT Asp	GAT Asp	TIT	GTA Val 660	Gly	GCT Ala	TTA Leu	ATA Ile	TTT Phe 665	TCA Ser	GGA Gly	GCT Ala	GTT Val	ATI Ile 670	Lev	TTA Leu	2016
30	Glu	Phe	Ile 675	Pro	Glu	Ile	Ala	Ile 680	Pro	Val	Leu	Gly	ACT Thr 685	Phe	Ala	<b>Le</b> u	2064
	Val	Ser 690	Tyr	Ile	Ala	Asn	Lys 695	Val	Leu	Thr	Val	Gln 700	ACA Thr	Ile	Asp	Asn	2112
35	Ala 705	Leu	Ser	Lys	Arg	Asn 710	Glu	Lys	Trp	Asp	Glu 715	Val	TAT Tyr	Lys	Tyr	Ile 720	2160
40	Val	Thr	Asn	Trp	<b>Leu</b> 725	Ala	Lys	Val	Asn	Thr 730	Gln	Ile	GAT Asp	Leu	11e 735	Arg	2208
	Lys	Lys	Met	Lys 740	Glu	Ala	Leu	Glu	Asn 745	Gln	Ala	Glu		Thr 750	Lys	Ala	2256
45		Ile	Asn 755	Tyr	Gln	Tyr	Asn	Gln 760	Tyr	Thr	Glu	Glu	Glu 765	Lys	Asn	Asn	2304
		Asn 770	Phe	Asn	Ile	Asp	Asp 775	Leu	Ser	Ser	Lys	Leu 780	Asn (	Glu	Ser	Ile	2352
50	AAT ASD 785	Lys	Ala	Met	Ile	<b>Asn</b> 790	Ile .	Asn :	Lys	Phe :	Leu . 795	Asn	Gln (	Cys	Ser	Val 800	2400
55	TCA :	IAT '	TTA Leu	Met .	AAT Asn 805	TCT : Ser !	ATG . Met	ATC (	Pro '	TAT ( Tyr ( 810	GGT (	GTT . Val 1	AAA ( Lys )	\rg	TTA Leu 815	GAA Glu	2448

5	GAT Asp	TTT	GAT Asp	GC1 Ala 820	Ser	CTI Leu	'AAA Lys	GAT Asp	GCA Ala 825	ı Lev	TTA Leu	AAC Lys	TAT	ATA Ile	Tyz	GAT Asp	2496
	AAT Asn	AGA Arg	GGA Gly 835	Thr	TTA	ATT	GGT Gly	CAA Gln 840	Val	GAT Asp	AGA Arg	TTA Leu	AAA Lys 845	Asp	Lys	GTT Val	2544
10	AAT Asn	AAT Asn 850	Thr	CTT Leu	AGT Ser	ACA Thr	GAT Asp 855	ATA Ile	CCT Pro	TTT Phe	CAG Gln	CTT Leu 860	TCC Ser	AAA Lys	TAC	GTA Val	2592
15	GAT Asp 865	AAT Asn	CAA Gln	AGA Arg	TTA Leu	TTA Leu 870	TCT Ser	ACA Thr	TTT Phe	ACT Thr	GAA Glu 875	TAT Tyr	ATT Ile	AAG Lys	TCT Ser	AGG Arg 880	2640
,3	CCT Pro	GGA Gly	CCG Pro	GAG Glu	ACG Thr 885	CTC Leu	TGC Cys	GGG Gly	GCT Ala	GAG Glu 890	CTG Leu	GTG Val	GAT Asp	GCT Ala	CTT Leu 895	CAG Gln	2688
20	TTC Phe	GTG Val	TGT Cys	GGA Gly 900	GAC Asp	AGG Arg	GGC Gly	TTT Phe	TAT Tyr 905	TTC Phe	AAC Asn	AAG Lys	CCC Pro	ACA Thr 910	GGG Gly	TAT Tyr	2736
	GGC Gly	TCC Ser	AGC Ser 915	AGT Ser	CGG Arg	AGG Arg	Ala	CCT Pro 920	CAG Gln	ACA Thr	GGT Gly	ATC Ile	GTG Val 925	GAT Asp	GAG Glu	TGC Cys	2784
25	Cys	TTC Phe 930	CGG Arg	AGC Ser	TGT Cys	Asp	CTA Leu 935	AGG Arg	AGG Arg	CTG Leu	Glu	ATG Met 940	TAT Tyr	TGC Cys	GCA Ala	CCC Pro	2832
30				GCC Ala	Lys												2862

- (2) INFORMATION FOR SEQ ID NO: 14:
- 35 (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 954 amino acids (B) TYPE: amino acid

  - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: protein
    - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

55

40

45

	Met 1	Gln	Phe	Val	Asn 5	Lys	Gln	Phe	Asn	Tyr 10	Lys	Asp	Pro	Va1	Asn 15	Gly
5	Val	Asp	Ile	Ala 20	туг	Ile	Lys	Ile	Pro 25	Asn	Ala	Gly	Gln	Met 30	Gln	Pro
	Val	Lys	Ala 35	Phe	Lys	Ile	His	Asn 40	Lys	Ile	Trp	Val	Ile 45	Pro	Glu	Arg
10	Asp	Thr 50	Phe	Thr	Asn	Pro	Glu 55	Glu	Gly	Asp	Leu	neA 00	Pro	Pro	Pro	Glu
	Ala 65	Lys	Gln	Val	Pro	Val 70	Ser	Tyr	Tyr	Ąsp	Ser 75	Thr	Tyr	Leu	Ser	Thr 80
15	Asp	Asn	Glu	Lys	Asp 85	Asn	Tyr	Leu	Lys	Gly 90	Val	Thr	Lys	Leu	Phe 95	Glu
	Arg	Ile	Tyr	Ser 100	Thr	Asp	Leu	Gly	Arg 105	Met	Leu	Leu	Thr	Ser 110	Ile	Val
20																
25																
30																

30
35
40
45

50

	Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys
5	Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 135 140
	Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160
10	Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175
	Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190
15	Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205
	Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 220
20	Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 230 240
	Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255
25	Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270  Phe Ile Asp Ser Leu Glo Glu Asp Cly Phe I
	Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn 275 280 285  Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290
30	Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Dhe Lys Clu Line
35	Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu
55	Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp
40	Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355 360 365
	Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 375 380
45	Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 385 390 395 400
	Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 415
50	Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg
	Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys 435 440 445
55	Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val 450 455 460

	As 46	sn A 55	sn :	rp	Asp	Le	1 Ph	e Ph O	e Se	r P	ro	Ser	Gl: 479	ı Ası	p As	n P	he T	hr	As:
5	As	p L	eu A	Asn	Lys	Gl) 485	/ Gli	u Gl	u Il	e T	hr	Ser 490	Asp	Th	r As	n I		lu 95	Ala
	Al	a G	lu G	lu	Asn 500	Ile	Se:	r Le	u As	p L	eu 05	Ile	Gln	Glr	з Ту	r Ty 51		eu	Thr
10	Ph	e A	sn P 5	he 15	Asp	Asn	Glu	ı Pro	52	u As O	n i	lle	Ser	Ile	G1 52		n Le	eu	Ser
	Se	r As 53	sp I	le	Ile	Gly	Gln	Le: 539	ı Gl	u Le	eu I	1et	Pro	Asn 540		e Gl	u Az	9	Phe
15	Pre 54!	o As 5	en G	1y :	Lys	Lys	Tyr <b>5</b> 50	Gli	Le	u As	p I	Lys	Tyr 555	Thr	Me	t Ph	e Hi		Tyr 560
	Let	ı Az	g A	la (	31n	Glu 565	Phe	Glu	Hi:	s Gl	y I 5	ys 70	Ser	Arg	Ile	<b>A</b> 1	a Le 57		fhr
20	Asr	ı Se	r V	al <i>)</i>	asn 80	Glu	Ala	Leu	Let	1 As 58	n P 5	ro	Ser	Arg	Va]	Ty:		r I	?he
	Phe	: Se	r Se	er A 95	dsk	Tyr	Val	Lys	Lys 600	Va	1 A	sn	Lys	Aļa	Thi 605		ı Al	a A	la
25	Met	Ph 61	e Le O	eu G	Sly '	Trp	Val	Glu 615	Gln	Le	u V	al	Tyr	Asp 620	Phe	Tha	: As	p G	lu
	Thr 625	Se:	r Gl	u V	al :	Ser	Thr 630	Thr	Asp	Lys	s I.		Ala 635	Asp	Ile	Thr	Ile		le 40
30					•	545		Ala			6	50					655	5	-
				6	60			Leu		665	;					670			
35			67	5				Ala	680						685				
		590	,					Lys 695					•	700					
40	Ala 705						710					7	15					72	20
	Val				7	25					73	0					735		_
45	Lys			74	0					745						750	-		
	Ile		755	•					760					7	765				
50		770					7	775					7	80					
	Asn 785					7	90					7	95					80	0
55	Ser '	Tyr	Leu	Me	E A:	sn S 05	er M	iet :	Ile	Pro	Ty:	c G)	ly V	al L	ys i		Leu 815	G1:	u

	A	.sp	Phe	Asp	Ala 820	Ser	Leu	Lys	Asp	Ala 825	Leu	Leu	Lys	Tyr	Ile 830	Tyr	Ası
5	A	sn	Arg	Gly 835	Thr	Leu	Ile	Gly	Gln 840	Val	Ąsp	Arg	Leu	Lys 845	Asp	Lys	Va 1
	A	sn	Asn 850	Thr	Leu	Ser	Thr	Asp 855	Ile	Pro	Phe	Gln	Leu 860	Ser	Lys	Tyr	Val
10	A:	sp 55	Asn	Gln	Arg	Leu	Leu 870	Ser	Thr	Phe	Thr	Glu 875	Tyr	Ile	Lys	Ser	Arg 880
	P	co	Gly	Pro	Glu	Thr 885	Leu	Сув	Gly	Ala	Glu 890	Leu	Val	Asp	Ala	Leu 895	Gln
15	Pl	10	Val	Cys	Gly 900	Asp	Arg	Gly	Phe	Tyr 905	Phe	Asn	Lys	Pro	Thr 910	Gly	Tyr
	G]	y	Ser	Ser 915	Ser	Arg	Arg	Ala	Pro 920	Gln	Thr	Gly	Ile	Val 925	Asp	Glu	Cys
20	Су	'S	Phe 930	Arg	Ser	Суз	Asp	Leu 935	Arg	Arg	Leu	Glu	Met 940	Tyr	Cys	Ala	Pro
	Le 94		Lys	Pro	Ala	Lys	Ser 950	Ala	Glu	Ala	*						
25	(2)	IN	FORM	MATIO	N FO	R SEC	ID N	O: 15:	:								
		(i	) SEQ	UENC	CE CH	ARAC	TERI	STICS	S:								
30			(B) (C)	TYPE STRA	: nucl	?724 b eic aci DNES: Y: line	id S: dou										
35		(ii	) MOL	.ECUl	E TY	PE: DI	VA (ge	enomi	c)								
		(i)	c) FEA	TURE	:												
40						: CDS :1272											
		(xi	) SEC	QUEN	CE DE	SCRI	PTION	1: SE(	N OI C	IO: 15	:						
45																	

	ATG Met 1	CAG Gln	TTC Phe	GTG Val	AAC Asn 5	AAG Lys	CAG Gln	TTC Phe	AAC Asn	TAT Tyr 10	AAG Lys	GAC Asp	CCT Pro	GTA Val	AAC Asn 15	GGT Gly	4.6
5	GTT Val	GAC Asp	ATT Ile	GCC Ala 20	TAC Tyr	ATC Ile	AAA Lys	ATT Ile	CCA Pro 25	AAC Asn	GCC Ala	GGC Gly	CAG Gln	ATG Met 30	CAG Gln	CCG Pro	96
10	GTG Val	AAG Lys	GCT Ala 35	TTC Phe	AAG Lys	ATT Ile	CAT His	AAC Asn 40	AAA Lys	ATC Ile	TGG Trp	GTT Val	ATT Ile 45	CCG Pro	GAA Glu	CGC Arg	144
	GAT Asp	ACA Thr 50	TTT Phe	ACG Thr	AAC Asn	CCG Pro	GAA Glu 55	GAA Glu	GGA Gly	GAC Asp	TTG Leu	AAC Asn 60	CCG Pro	CCG Pro	CCG Pro	GAA Glu	192
15	GCA Ala 65	AAG Lys	CAG Gln	GTG Val	CCA Pro	GTT Val 70	TCA Ser	TAC Tyr	TAC Tyr	GAT Asp	TCA Ser 75	ACC Thr	TAT Tyr	CTG Leu	AGC Ser	ACA Thr 80	240
20	GAC Asp	AAC Asn	GAG Glu	AAG Lys	GAT Asp 85	AAC Asn	TAC Tyr	CTG Leu	AAG Lys	GGA Gly 90	GTG Val	ACC Thr	AAA Lys	TTA Leu	TTC Phe 95	GAG Glu	288

	CGT ATT TAT TCC ACT GAC CTG GGC CGT ATG CTG CTG ACC TCA ATC GTC Arg lle Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val	336
5	CGC GGA ATC CCA TTT TGG GGT GGC AGT ACC ATT GAC ACG GAG TTG AAG Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys	384
10	GTT ATT GAC ACT AAC TGC ATT AAC GTG ATC CAA CCA GAC GGT AGC TAC Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130	432
4.5	AGA TCT GAA GAA CTT AAC CTC GTA ATC ATC GGG CCC TCC GCG GAC ATT Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 150 155 160	480
15	ATC CAG TTT GAG TGC AAG AGC TTT GGC CAC GAA GTG TTG AAC CTG ACG Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 170 CCT AAC GCT TAG GCT TAG AGC TTT GGC CAC GAA GTG TTG AAC CTG ACG 165 175	528
20	CGT AAC GGT TAC GGC TCT ACT CAG TAC ATT CGT TTC AGC CCA GAC TTC Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 185	576
	ACG TTC GGT TTC GAG GAG AGC CTG GAG GTT GAT ACC AAC CCG CTG TTG Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 200 205	624
25	GGT GCA GGC AAG TTC GCA ACT GAT CCA GCG GTG ACC CTG GCA CAC GAG Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 215 220	672
30	CTG ATC CAC GCC GGT CAT CGT CTG TAT GGC ATT GCG ATT AAC CCG AAC Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 230	720
	CGC GTG TTC AAG GTT AAC ACC AAC GCC TAC TAC GAG ATG AGT GGT TTA Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245	768
35	GAA GTA AGC TTC GAG GAA CTG CGC ACG TTC GGT GGC CAT GAT GCG AAG Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265	816
40	TTT ATC GAC AGC TTG CAG GAG AAC GAG TTC CGT CTG TAC TAC TAC AAC Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn 285 285	864
	AAG TIT AAA GAT ATT GCA AGT ACA CTG AAC AAG GCT AAG TCC ATT GTG Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300	912
<b>4</b> 5	GGT ACC ACT GCT TCA TTA CAG TAT ATG AAA AAT GTT TTT AAA GAG AAA GIY Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 310 315 320	960
	TAT CTC CTA TCT GAA GAT ACA TCT GGA AAA TTT TCG GTA GAT AAA TTA Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 330 335	1008
50	AAA TTT GAT AAG TTA TAC AAA ATG TTA ACA GAG ATT TAC ACA GAG GAT Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340 345	1056
55	AAT TIT GIT AAG TIT TIT AAA GTA CIT AAC AGA AAA ACA TAT TIG AAT Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355	1104

5	TTT GAT AAA GCC GTA TTT AAG ATA AAT ATA GTA CCT AAG GTA AAT TAC Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 375	1152
5	ACA ATA TAT GAT GGA TTT AAT TTA AGA AAT ACA AAT TTA GCA GCA AAC Thr lle Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 385 390 395	1200
10	TTT AAT GGT CAA AAT ACA GAA ATT AAT AAT ATG AAT TTT ACT AAA CTA Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 415	1248
	AAA AAT TIT ACT GGA TTG TIT GAA TTT TAT AAG TTG CTA TGT GTA AGA Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 420 425 430	1296
15	GGG ATA ATA ACT TCT AAA ACT AAA TCA TTA GAT AAA GGA TAC AAT AAG Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys 435 440 445	1344
20	ATC GAA GGT CGT TGC GAT GGG GCA TTA AAT GAT TTA TGT ATC AAA GTT Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val 450 460	1392
	AAT AAT TGG GAC TTG TTT TTT AGT CCT TCA GAA GAT AAT TTT ACT AAT Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 465 470 480	1440
25	GAT CTA AAT AAA GGA GAA GAA ATT ACA TCT GAT ACT AAT ATA GAA GCA Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 485 490 495	1488
30	GCA GAA GAA AAT ATT AGT TTA GAT TTA ATA CAA CAA TAT TAT TTA ACC Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 500 500	1536
	TTT AAT TIT GAT AAT GAA CCT GAA AAT ATT TCA ATA GAA AAT CTT TCA Phe Asn Phe Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 515 520 525	1584
35	AGT GAC ATT ATA GGC CAA TTA GAA CTT ATG CCT AAT ATA GAA AGA TTT Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe 535 540	1632
40	CCT AAT GGA AAA AAG TAT GAG TTA GAT AAA TAT ACT ATG TTC CAT TAT Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr 545 555 560	1680
	CTT CGT GCT CAA GAA TTT GAA CAT GGT AAA TCT AGG ATT GCT TTA ACA Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr 565 570 575	1728
45	AAT TCT GTT AAC GAA GCA TTA TTA AAT CCT AGT CGT GTT TAT ACA TTT Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe 580 585 590	1776
	TTT TCT TCA GAC TAT GTA AAG AAA GTT AAT AAA GCT ACG GAG GCA GCT Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala 595 600 605	1824
50	ATG TTT TTA GGC TGG GTA GAA CAA TTA GTA TAT GAT TTT ACC GAT GAA Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu 610 620	1872
55	ACT AGC GAA GTA AGT ACT ACG GAT AAA ATT GCG GAT ATA ACT ATA ATT Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 625 630 640	1920

	ATT CCA TAT ATA GGA CCT GCT TTA AAT ATA GGT AAT ATG TTA TAT AAA Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys 645 650	
5	GAT GAT TIT GTA GGT GCT TTA ATA TIT TCA GGA GCT GTT ATT CTG TTA Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 660 665	2016
10	GAA TIT ATA CCA GAG ATT GCA ATA CCT GTA TTA GGT ACT TIT GCA CTT Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 675 680 685	2064
	GTA TCA TAT ATT GCG AAT AAG GTT CTA ACC GTT CAA ACA ATA GAT AAT Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 690 700	2112
15	GCT TTA AGT AAA AGA AAT GAA AAA TGG GAT GAG GTC TAT AAA TAT ATA Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile 715	2160
20	GTA ACA AAT TGG TTA GCA AAG GTT AAT ACA CAG ATT GAT CTA ATA AGA Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg 735 736 737	2208
	AAA AAA ATG AAA GAA GCT TTA GAA AAT CAA GCA GAA GCA ACA AAG GCT Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala 740 745 750	2256
25	ATA ATA AAC TAT CAG TAT AAT CAA TAT ACT GAG GAA GAG AAA AAT AAT Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn 755 760 765	2304
30	ATT AAT TIT AAT ATT GAT GAT TTA AGT TCG AAA CTT AAT GAG TCT ATA Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile 775 780	2352
	AAT AAA GCT ATG ATT AAT ATA AAT AAA TTT TTG AAT CAA TGC TCT GTT Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 795 800	2400
35	TCA TAT TTA ATG AAT TCT ATG ATC CCT TAT GGT GTT AAA CGG TTA GAA Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 805 810 815	2448
40	GAT TIT GAT GCT AGT CTT AAA GAT GCA TTA TTA AAG TAT ATA TAT GAT ASP Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp 820 825 830	2496
40	AAT AGA GGA ACT TTA ATT GGT CAA GTA GAT AGA TTA AAA GAT AAA GTT Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val 835	2544
45	AAT AAT ACA CTT AGT ACA GAT ATA CCT TTT CAG CTT TCC AAA TAC GTA Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val 850	2592
	GAT AAT CAA AGA TTA TTA TCT ACA TTT ACT GAA TAT ATT AAG TCT AGG Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys Ser Arg 870 875 880	2640
50	CCT CAA TCT AAA GTT AAA AGA CAA ATA TTT TCA GGC TAT CAA TCT GAT Pro Gln Ser Lys Val Lys Arg Gln ile Phe Ser Gly Tyr Gln Ser Asp 890 895	2688
55	ATT GAT ACA CAT AAT AGA ATT AAG GAT GAA TTA TGA Ile Asp Thr His Asn Arg Ile Lys Asp Glu Leu * 900 905	2724

	EP 0 9	339
	(2) INFORMATION FOR SEQ ID NO: 16:	
	(i) SEQUENCE CHARACTERISTICS:	
5	<ul><li>(A) LENGTH: 908 amino acids</li><li>(B) TYPE: amino acid</li><li>(D) TOPOLOGY: linear</li></ul>	
10	(ii) MOLECULE TYPE: protein	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1	6:
15		
20		
25		
30		
35		

	Moto (1) a table to a
	Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly 1 5 10 15
5	Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro 20 25 30
	Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg 35 40 45
10	Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 60
	Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80
15	Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu
	Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val
20	Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys 115 120 125
	Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 135 140
25	Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160
	Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170
30	Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190
	Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205
35	Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 225 220
	Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 230 235 240
<b>4</b> 0	Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255
	Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270
45	Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn 275 280 285
	Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300

	Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys
5	Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 330 335
	Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340 345 350
10	Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355 360 365
	Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 380
15	Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 385 390 395 400
	Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410
20	Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 420 425 430
	Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys 435 440 445
25	Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val 450 460
	Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 475 470 480
30	Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 485 490 495
-	Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 500 505 510  Phe Asn Phe Asn Asn Clu Dro Gl
35	Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 515 520 525  Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe 530 535
40	Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr  530  540  Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr  545
40	545 550 550 555 560 Leu Arg Ala Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr
<b>4</b> 5	Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe
	Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala
50	Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Clu
	Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile
55	Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys
	650 655

											00.	_					67	0	eu Lev
5																685	5		la Leu
	Va	al S	er 3	yr	Ile	Al	a As	n L	y <b>s</b> 95	/al	Lei	ı Th	r Va	1 G	ln 00	Thi	: I1	e As	p Asn
10							_	-					/1	3					r Ile 720
46												131	U					73	
15											, 13						750	)	s Ala
00									•						7	65			a Asn
20									_					78	G.				Ile
													793						Val 800
25												OIU						815	Glu
										٠	23						830		Asp
30										•			Arg		84	5			
													Gln	990					
35	Asp 865						•						0/5						880
	Pro					-						90		Gly	ту	r G	ln :	Ser 895	Asp
40	Ile	Asp	Thr	Hi 90	s A: 0	n A	Arg	Ile	Lys	As 90	sp G S	lu 1	Leu	•					

# (2) INFORMATION FOR SEQ ID NO: 17:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 3042 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (ix) FEATURE:

45

50

- (A) NAME/KEY: CDS (B) LOCATION:1..3042

	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:																	
5	ATG Met 1	CAG Gln	TTC Phe	GTG Val	AAC Asn 5	AAG Lys	CAG Gln	TTC Phe	AAC Asn	TAT Tyr 10	AAG Lys	GAC Asp	CCT Pro	GTA Val	AAC Asn 15	GGT Gly		49
10																		
15																		
20																		
25																		
30																		

	GTT GAC ATT GCC TAC ATC AAA ATT CCA AAC GCC GGC CAG ATG CAG CCG Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro 20 25 30	96
5	GTG AAG GCT TTC AAG ATT CAT AAC AAA ATC TGG GTT ATT CCG GAA CGC Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg	144
10	GAT ACA TTT ACG AAC CCG GAA GAA GGA GAC TTG AAC CCG CCG CCG GAA Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 60	192
45	GCA AAG CAG GTG CCA GTT TCA TAC TAC GAT TCA ACC TAT CTG AGC ACA Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 70 75	240
15	GAC AAC GAG AAG GAT AAC TAC CTG AAG GGA GTG ACC AAA TTA TTC GAG Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90	288
20	CGT ATT TAT TCC ACT GAC CTG GGC CGT ATG CTG CTG ACC TCA ATC GTC ACG Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val	336
	CGC GGA ATC CCA TTT TGG GGT GGC AGT ACC ATT GAC ACG GAG TTG AAG Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys 120 GTT ATT GAC ACT 110 TGG	384
25	GTT ATT GAC ACT AAC TGC ATT AAC GTG ATC CAA CCA GAC GGT AGC TAC Val lle Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 135	432
30	AGA TCT GAA GAA CTT AAC CTC GTA ATC ATC GGG CCC TCC GCG GAC ATT Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 150 150 160	480
	ATC CAG TTT GAG TGC AAG AGC TTT GGC CAC GAA GTG TTG AAC CTG ACG Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 170 175	528
35	CGT AAC GGT TAC GGC TCT ACT CAG TAC ATT CGT TTC AGC CCA GAC TTC AGG Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 185	576
<b>4</b> 0	ACG TTC GGT TTC GAG GAG AGC CTG GAG GTT GAT ACC AAC CCG CTG TTG Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 200 205	624
	GGT GCA GGC AAG TTC GCA ACT GAT CCA GCG GTG ACC CTG GCA CAC GAG Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 215 220	672
<b>4</b> 5	CTG ATC CAC GCC GGT CAT CGT CTG TAT GGC ATT GCG ATT AAC CCG AAC Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 235 240	720
50	CGC GTG TTC AAG GTT AAC ACC AAC GCC TAC TAC GAG ATG AGT GGT TTA Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255	768
	GAA GTA AGC TTC GAG GAA CTG CGC ACG TTC GGT GGC CAT GAT GCG AAG Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270	816
55	TTT ATC GAC AGC TTG CAG GAG AAC GAG TTC CGT CTG TAC TAC TAC AAC Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn 285	864

5	AAG TTT AAA GAT ATT GCA AGT ACA CTG AAC AAG GCT AAG TCC ATT GTG Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300	912
J	GGT ACC ACT GCT TCA TTA CAG TAT ATG AAA AAT GTT TTT AAA GAG AAA Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 310 315 320	960
10	TAT CTC CTA TCT GAA GAT ACA TCT GGA AAA TTT TCG GTA GAT AAA TTA Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 330 335	1008
15	AAA TTT GAT AAG TTA TAC AAA ATG TTA ACA GAG ATT TAC ACA GAG GAT Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340 345 350	1056
	AAT TTT GTT AAG TTT TTT AAA GTA CTT AAC AGA AAA ACA TAT TTG AAT Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355 360 365	1104
20	TTT GAT AAA GCC GTA TTT AAG ATA AAT ATA GTA CCT AAG GTA AAT TAC Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 375 380	1152
25	ACA ATA TAT GAT GGA TTT AAT TTA AGA AAT ACA AAT TTA GCA GCA AAC Thr lle Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 385 390 400	1200
20	TTT AAT GGT CAA AAT ACA GAA ATT AAT AAT ATG AAT TTT ACT AAA CTA Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 410 415	1248
30	AAA AAT TTT ACT GGA TTG TTT GAA TTT TAT AAG TTG CTA TGT GTA AGA Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 420 425 426	1296
	GGG ATA ATA ACT TCT AAA ACT AAA TCA TTA GAT AAA GGA TAC AAT AAG Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys 445	1344
35	ATC GAA GGT CGT TGC GAT GGG GCA TTA AAT GAT TTA TGT ATC AAA GTT 11e Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val 450 451	1392
40	AAT AAT TGG GAC TTG TTT TTT AGT CCT TCA GAA GAT AAT TTT ACT AAT Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 470 480	1440
	GAT CTA AAT AAA GGA GAA GAA ATT ACA TCT GAT ACT AAT ATA GAA GCA Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 485 490 495	1488
<b>4</b> 5	GCA GAA GAA AAT ATT AGT TTA GAT TTA ATA CAA CAA TAT TAT TTA ACC Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 500 510	1536
50	TTT AAT TTT GAT AAT GAA CCT GAA AAT ATT TCA ATA GAA AAT CTT TCA Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 515 520 525	1584
	530 535 S40 Silv Bed Met Pro Asn Ile Glu Arg Phe	1632
55	CCT AAT GGA AAA AAG TAT GAG TTA GAT AAA TAT ACT ATG TTC CAT TAT Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr 545 550 560	1680

CGT	GCT	CAA	GAA	TTT	GAA	CAT	GGT	222	7

5	CTT CGT GCT CAA GAA TTT GAA CAT GGT AAA TCT AGG ATT GCT TTA ACA Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr 565 570	1728
	AAT TCT GTT AAC GAA GCA TTA TTA AAT CCT AGT CGT GTT TAT ACA TTT Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe 585	1776
10	TTT TCT TCA GAC TAT GTA AAG AAA GTT AAT AAA GCT ACG GAG GCA GCT Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala 595 600 605	1824
15	ATG TTT TTA GGC TGG GTA GAA CAA TTA GTA TAT GAT TTT ACC GAT GAA Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu 610 620	1872
	ACT AGC GAA GTA AGT ACT ACG GAT AAA ATT GCG GAT ATA ACT ATA ATT Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 635 640 ATT CCA TAT ATA CCA CCT CCT TO	1920
20	ATT CCA TAT ATA GGA CCT GCT TTA AAT ATA GGT AAT ATG TTA TAT AAA  1le Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys 645 655 GAT GAT TTT GTA GGT CCT TTA ATA TTT	1968
25	GAT GAT TIT GTA GGT GCT TTA ATA TIT TCA GGA GCT GTT ATT CTG TTA ASP ASP Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 660 665 GAA TIT ATA CCA GAG ATT GCA ATA CCT GTA TTA GGT ACT TIT GCA CTT Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly ACT TIT GCA CTT	2016
	675 680 685	2064
30	GTA TCA TAT ATT GCG AAT AAG GTT CTA ACC GTT CAA ACA ATA GAT AAT Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gin Thr Ile Asp Asn 690 695 700	2112
	GCT TTA AGT AAA AGA AAT GAA AAA TGG GAT GAG GTC TAT AAA TAT ATA Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile 715 710 720	2160
35	GTA ACA AAT TGG TTA GCA AAG GTT AAT ACA CAG ATT GAT CTA ATA AGA Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg 725 730 735	2208
40	AAA AAA ATG AAA GAA GCT TTA GAA AAT CAA GCA GAA GCA ACA AAG GCT Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala 745 750	2256
	ATA ATA AAC TAT CAG TAT AAT CAA TAT ACT GAG GAA GAG AAA AAT AAT Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn 755 760 765	2304
45	ATT AAT TIT AAT ATT GAT GAT TTA AGT TCG AAA CTT AAT GAG TCT ATA Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile 770 780	2352
50	AAT AAA GCT ATG ATT AAT ATA AAT AAA TIT TIG AAT CAA TGC TCT GTT Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 785 790 800	2400
	TCA TAT TTA ATG AAT TCT ATG ATC CCT TAT GGT GTT AAA CGG TTA GAA Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 805 815	2448
55	GAT TIT GAT GCT AGT CIT AAA GAT GCA TTA TTA AAG TAT ATA TAT GAT Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Lys Tyr Ile Tyr Asp 820 825	2496

5	AAT AGA GGA ACT TTA ATT GGT CAA GTA GAT AGA TTA AAA GAT AAA GTT Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val 835 840 845	2544
	AAT AAT ACA CTT AGT ACA GAT ATA CCT TTT CAG CTT TCC AAA TAC GTA Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val 850	2592
10	GAT AAT CAA AGA TTA TTA TCT ACA TTT ACT GAA TAT ATT AAG TCA GGC Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys Ser Gly 875	2640
15	CTG AAT TCC CCG GGT GCA GCT CAT TAT GCG CAA CAC GAT GAA GCC GTA Leu Asn Ser Pro Gly Ala Ala His Tyr Ala Gln His Asp Glu Ala 885	2688
	GAC AAC AAA TTC AAC AAA GAA CAA CAA AAC GCG TTC TAT GAG ATC TTA Asp Asn Lys Phe Asn Lys Glu Gln Asn Ala Phe Tyr Glu Ile Leu 900 905 910	2736
20	CAT TTA CCT AAC TTA AAC GAA GAA CAA CGA AAC GCC TTC ATC CAA AGT His Leu Pro Asn Leu Asn Glu Glu Gln Arg Asn Ala Phe Ile Gln Ser 915 920 925	2784
	TTA AAA GAT GAC CCA AGC CAA AGC GCT AAC CTT TTA GCA GAA GCT AAA Leu Lys Asp Asp Pro Ser Gln Ser Ala Asn Leu Leu Ala Glu Ala Lys 930 935	2832
25	AAG CTA AAT GAT GCT CAG GCG CCG AAA GTA GAC AAC AAA TTC AAC AAA Lys Leu Asn Asp Ala Gln Ala Pro Lys Val Asp Asn Lys Phe Asn Lys 945 950 955	2880
30	GAA CAA CAA AAC GCG TTC TAT GAG ATC TTA CAT TTA CCT AAC TTA AAC Glu Gln Gln Asn Ala Phe Tyr Glu Ile Leu His Leu Pro Asn Leu Asn 965 970 975	2928
	GAA GAA CAA CGA AAC GCC TTC ATC CAA AGT TTA AAA GAT GAC CCA AGC Glu Glu Gln Arg Asn Ala Phe Ile Gln Ser Leu Lys Asp Asp Pro Ser 980 985 990	2976
35	CAA AGC GCT AAC CTT TTA GCA GAA GCT AAA AAG CTA AAT GAT GCT CAG Gln Ser Ala Asn Leu Leu Ala Glu Ala Lys Lys Leu Asn Asp Ala Gln 995 1000 1005	3024
40	GCG CCG AAA GTA GAC TAG Ala Pro Lys Val Asp * 1010	3042

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1014 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

55

45

	-: 0 000 010 B1															
															Asn 15	
5															Gln	
	Val	Lys	Ala 35	Phe	Lys	Ile	His	Asn 40	Lys	Ile	Trp	Val	Ile 45	Pro	Glu	Arg
10																
15																
20																
25																
30																

	Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu
5	Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr  70  75
	Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu
10	Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Vai
45	Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys 115 120 125 Val Ile Asp Thr Asp Grant
15	Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 135 140 Arg Ser Glu Glu Leu Ass I
20	Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160
20	Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175  Arg Asn Gly Tyr Gly Ser The Gla Tarana
25	Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190 Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195
	Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu
30	Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 230
	Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250
35	Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265
	Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn 275 280
40	Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300
	Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 305 310 315 320
45	Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 330 335 Lys Phe Asp Lys Leu Tyr Lys Mar I
	Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340 345 350  Asn Phe Val Lys Phe Phe Lys Val Lau
50	Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 360 365  Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 375
	Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn
55	390 395 Leu Ala Ala Asn 400

	Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410
5	Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 420 425 430
	Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys 435 440 445
10	Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val 450 460
	Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 465 470 480
15	Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 485 490 495
	Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 500 505 510
20	Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 515 520 525
	Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe 530 540
25	Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr 545 550 555 560
	Leu Arg Ala Glu Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr 565 570 575
30	Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe 580 590
	Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala 595 600 605
35	Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu 610 615 620  Thr Ser Glu Val Ser Thr Thr Asp Jun Th
	Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 625 630 635 640  Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys
40	Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu  645  Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu
	Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 675
45	Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn
	Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile
50	710 715 720  Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg
	The Asp Leu Ile Arg 730  Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala 740  745
55	740 745 750

	Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn 755 760 765
5	Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile 770 780
	Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val
10	Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 805 810 815
	Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp 820 825 830
15	Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val 835 840 845
	Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val 850 855 860
20	Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys Ser Gly 865 870 875 880
	Leu Asn Ser Pro Gly Ala Ala His Tyr Ala Gln His Asp Glu Ala Val 885 890 895
25	Asp Asn Lys Phe Asn Lys Glu Gln Asn Ala Phe Tyr Glu Ile Leu 900 910
	His Leu Pro Asn Leu Asn Glu Glu Gln Arg Asn Ala Phe Ile Gln Ser 915 920 925
30	Leu Lys Asp Asp Pro Ser Gln Ser Ala Asn Leu Leu Ala Glu Ala Lys 930 935 940
	Lys Leu Asn Asp Ala Gln Ala Pro Lys Val Asp Asn Lys Phe Asn Lys 945 950 955 960
35	Glu Gln Gln Asn Ala Phe Tyr Glu Ile Leu His Leu Pro Asn Leu Asn 965 970 975
	Glu Glu Gln Arg Asn Ala Phe Ile Gln Ser Leu Lys Asp Asp Pro Ser 985 990
40	Gln Ser Ala Asn Leu Leu Ala Glu Ala Lys Lys Leu Asn Asp Ala Gln 995 1000 1005
	Ala Pro Lys Val Asp + 1010
45	(2) INFORMATION FOR SEQ ID NO: 19:
	(i) SEQUENCE CHARACTERISTICS:
50	<ul><li>(A) LENGTH: 3509 base pairs</li><li>(B) TYPE: nucleic acid</li><li>(C) STRANDEDNESS: double</li><li>(D) TOPOLOGY: linear</li></ul>
55	(ii) MOLECULE TYPE: DNA (genomic)
	(ix) FEATURE:
	(A) NAME/KEY: CDS

(B) LOCATION:1..3509

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

5	ATG CCA GTT ACA ATA AAT AAT TTT AAT TAT AAT GAT CCT ATT GAT AAT Met Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile Asp Asn 1 5 10	48
3	AAT AAT ATT ATG ATG GAG CCT CCA TTT GCG AGA GGT ACG GGG AGA Asn Asn Ile Ile Met Met Glu Pro Pro Phe Ala Arg Gly Thr Gly Arg 20 25 30	96
10	TAT TAT AAA GCT TTT AAA ATC ACA GAT CGT ATT TGG ATA ATA CCG GAA Tyr Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile Pro Glu 35	144
15	AGA TAT ACT TTT GGA TAT AAA CCT GAG GAT TTT AAT AAA AGT TCC GGT Arg Tyr Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser Gly 50 60	192
7.5	ATT TIT AAT AGA GAT GTT TGT GAA TAT TAT GAT CCA GAT TAC TTA AAT  1le Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp Tyr Leu Asn  65 70 75 80	240
20	ACT AAT GAT AAA AAG AAT ATA TTT TTA CAA ACA ATG ATC AAG TTA TTT Thr Asn Asp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile Lys Leu Phe 85 90 95	288
	AAT AGA ATC AAA TCA AAA CCA TTG GGT GAA AAG TTA TTA GAG ATG ATT Asn Arg Ile Lys Ser Lys Pro Leu Gly Glu Lys Leu Leu Glu Met Ile 100 105 110	336
25	ATA AAT GGT ATA CCT TAT CTT GGA GAT AGA CGT GTT CCA CTC GAA GAG Ile Asn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro Leu Glu Glu 115	384
30	TTT AAC ACA AAC ATT GCT AGT GTA ACT GTT AAT AAA TTA ATC AGT AAT Phe Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile Ser Asn 130 135	432
	CCA GGA GAA GTG GAG CGA AAA AAA GGT ATT TTC GCA AAT TTA ATA ATA Pro Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn Leu Ile Ile 150 150	480
35	TTT GGA CCT GGG CCA GTT TTA AAT GAA AAT GAG ACT ATA GAT ATA GGT Phe Gly Pro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile Asp Ile Gly 170 175	528
40	ATA CAA AAT CAT TTT GCA TCA AGG GAA GGC TTC GGG GGT ATA ATG CAA Ile Gln Asn His Phe Ala Ser Arg Glu Gly Phe Gly Gly Ile Met Gln 180 185	576
	ATG AAG TIT TGC CCA GAA TAT GTA AGC GTA TTT AAT AAT GTI CAA GAA Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Gln Glu 195 200 205	624
<b>4</b> 5	AAC AAA GGC GCA AGT ATA TTT AAT AGA CGT GGA TAT TTT TCA GAT CCA Asn Lys Gly Ala Ser Ile Phe Asn Arg Arg Gly Tyr Phe Ser Asp Pro 210 220	672
5 <i>0</i>	GCC TTG ATA TTA ATG CAT GAA CTT ATA CAT GTT TTA CAT GGA TTA TAT Ala Leu Ile Leu Met His Glu Leu Ile His Val Leu His Gly Leu Tyr 230 235 240	720
	GGC ATT AAA GTA GAT GAT TTA CCA ATT GTA CCA AAT GAA AAA AAA TTT Gly Ile Lys Val Asp Asp Leu Pro Ile Val Pro Asn Glu Lys Lys Phe 245 250 255	768
55	TTT ATG CAA TCT ACA GAT GCT ATA CAG GCA GAA GAA CTA TAT ACA TTT Phe Met Gin Ser Thr Asp Ala Ile Gin Ala Glu Glu Leu Tyr Thr Phe 260 265	816

	GGA GGA CAA GAT CCC AGC ATC ATA ACT CCT TCT ACG GAT AAA AGT ATC Gly Gly Gln Asp Pro Ser Ile Ile Thr Pro Ser Thr Asp Lys Ser Ile 285	864
5	TAT GAT AAA GTT TTG CAA AAT TTT AGA GGG ATA GTT GAT AGA CTT AAC Tyr Asp Lys Val Leu Gln Asn Phe Arg Gly Ile Val Asp Arg Leu Asn 290 295 300	912
10	AAG GTT TTA GTT TGC ATA TCA GAT CCT AAC ATT AAT ATT AAT ATA TAT Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr 310 310 320	960
15	AAA AAT AAA TTT AAA GAT AAA TAT AAA TTC GTT GAA GAT TCT GAG GGA Lys Asn Lys Phe Lys Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly 325 330 335	1008
-	AAA TAT AGT ATA GAT GTA GAA AGT TTT GAT AAA TTA TAT AAA AGC TTA Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu 340 345	1056
20	ATG TTT GGT TTT ACA GAA ACT AAT ATA GCA GAA AAT TAT AAA ATA AAA Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys 355 360 365	1104
	ACT AGA GCT TCT TAT TTT AGT GAT TCC TTA CCA CCA GTA AAA ATA AAA Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys 370 375 380	1152
25	AAT TTA TTA GAT AAT GAA ATC TAT ACT ATA GAG GAA GGG TTT AAT ATA Asn Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile 390 395 400	1200
30	TCT GAT AAA GAT ATG GAA AAA GAA TAT AGA GGT CAG AAT AAA GCT ATA Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile 405	1248
	AAT AAA CAA GCT TAT GAA GAA ATT AGC AAG GAG CAT TTG GCT GTA TAT ABN Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr 420 425 430	1296
35	AAG ATA CAA ATG TGT AAA AGT GTT AAA GCT CCA GGA ATA TGT ATT GAT Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp 435	1344
40	GTT GAT AAT GAA GAT TTG TTC TTT ATA GCT GAT AAA AAT AGT TTT TCA Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser 450	1392
	GAT GAT TTA TCT AAA AAC GAA AGA ATA GAA TAT AAT ACA CAG AGT AAT Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn 470 475 480	1440
45	TAT ATA GAA AAT GAC TTC CCT ATA AAT GAA TTA ATT TTA GAT ACT GAT Tyr lle Glu Asn Asp Phe Pro lle Asn Glu Leu lle Leu Asp Thr Asp 485 490 495	1488
	TTA ATA AGT AAA ATA GAA TTA CCA AGT GAA AAT ACA GAA TCA CTT ACT Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr 500 505 510	1536
50	GAT TTT AAT GTA GAT GTT CCA GTA TAT GAA AAA CAA CCC GCT ATA AAA Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys 515	1584
55	AAA ATT TTT ACA GAT GAA AAT ACC ATC TTT CAA TAT TTA TAC TCT CAG Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln 530 535	1632

5	ACA TTT CCT CTA GAT ATA AGA GAT ATA AGT TTA ACA TCT TCA TTT GAT Thr Phe Pro Leu Asp Ile Arg Asp Ile Ser Leu Thr Ser Ser Phe Asp 545 550 560	1680
ū	GAT GCA TTA TTA TTT TCT AAC AAA GTT TAT TCA TTT TCT ATG GAT Asp Ala Leu Leu Phe Ser Asn Lys Val Tyr Ser Phe Phe Ser Met Asp 565 570	1728
10	TAT ATT AAA ACT GCT AAT AAA GTG GTA GAA GCA GGA TTA TTT GCA GGT Tyr Ile Lys Thr Ala Asn Lys Val Val Glu Ala Gly Leu Phe Ala Gly 580 585	1776
15	TGG GTG AAA CAG ATA GTA AAT GAT TTT GTA ATC GAA GCT AAT AAA AGC Trp Val Lys Gln Ile Val Asn Asp Phe Val Ile Glu Ala Asn Lys Ser 600 605	1824
	AAT ACT ATG GAT AAA ATT GCA GAT ATA TCT CTA ATT GTT CCT TAT ATA Asn Thr Met Asp Lys Ile Ala Asp Ile Ser Leu Ile Val Pro Tyr Ile 610 620	1872
20	GGA TTA GCT TTA AAT GTA GGA AAT GAA ACA GCT AAA GGA AAT TTT GAA Gly Leu Ala Leu Asn Val Gly Asn Glu Thr Ala Lys Gly Asn Phe Glu 630 635 640	1920
	AAT GCT TTT GAG ATT GCA GGA GCC AGT ATT CTA CTA GAA TTT ATA CCA Asn Ala Phe Glu Ile Ala Gly Ala Ser Ile Leu Leu Glu Phe Ile Pro 645 650 655	1968
25	GAA CTT TTA ATA CCT GTA GTT GGA GCC TTT TTA TTA GAA TCA TAT ATT Glu Leu Leu Ile Pro Val Val Gly Ala Phe Leu Leu Glu Ser Tyr Ile 660 665	2016
30	GAC AAT AAA AAT AAA ATT ATT AAA ACA ATA GAT AAT GCT TTA ACT AAA Asp Asn Lys Asn Lys Ile Ile Lys Thr Ile Asp Asn Ala Leu Thr Lys 675 680 685	2064
	AGA AAT GAA AAA TGG AGT GAT ATG TAC GGA TTA ATA GTA GCG CAA TGG Arg Asn Glu Lys Trp Ser Asp Met Tyr Gly Leu Ile Val Ala Gln Trp 690 700	2112
35	CTC TCA ACA GTT AAT ACT CAA TTT TAT ACA ATA AAA GAG GGA ATG TAT Leu Ser Thr Val Asn Thr Gln Phe Tyr Thr Ile Lys Glu Gly Met Tyr 715 720	2160
40	AAG GCT TTA AAT TAT CAA GCA CAA GCA TTG GAA GAA ATA ATA AAA TAC Lys Ala Leu Asn Tyr Gln Ala Gln Ala Leu Glu Glu Ile Ile Lys Tyr 725 730 735	2208
	AGA TAT AAT ATA TAT TCT GAA AAA GAA AAG TCA AAT ATT AAC ATC GAT Arg Tyr Asn Ile Tyr Ser Glu Lys Glu Lys Ser Asn Ile Asn Ile Asp 740 750	2256
45	TTT AAT GAT ATA AAT TCT AAA CTT AAT GAG GGT ATT AAC CAA GCT ATA Phe Asn Asp Ile Asn Ser Lys Leu Asn Glu Gly Ile Asn Gln Ala Ile 755 760 765	2304
50	GAT AAT ATA AAT TIT ATA AAT GGA TGT TCT GTA TCA TAT TTA ATG Asp Asn Ile Asn Asn Phe Ile Asn Gly Cys Ser Val Ser Tyr Leu Met 770 780	2352
	AAA AAA ATG ATT CCA TTA GCT GTA GAA AAA TTA CTA GAC TTT GAT AAT Lys Lys Met Ile Pro Leu Ala Val Glu Lys Leu Leu Asp Phe Asp Asn 785 790 795	2400
55	ACT CTC AAA AAA AAT TTG TTA AAT TAT ATA GAT GAA AAT AAA TTA TAT Thr Leu Lys Lys Asn Leu Leu Asn Tyr Ile Asp Glu Asn Lys Leu Tyr 805 810 815	2448

5	TTG ATT GGA AGT GCA GAA TAT GAA AAA TCA AAA GTA AAT AAA TAC TTG Leu Ile Gly Ser Ala Glu Tyr Glu Lys Ser Lys Val Asn Lys Tyr Leu 820 825	2496
v	AAA ACC ATT AIG CCG TTT GAT CTT TCA ATA TAT ACC AAT GAT ACA ATA Lys Thr lle Met Pro Phe Asp Leu Ser lle Tyr Thr Asn Asp Thr lle 835 840 845	2544
10	CTA ATA GAA ATG TTT AAT AAA TAT AAT AGC GAA ATT TTA AAT AAT ATT Leu Ile Glu Met Phe Asn Lys Tyr Asn Ser Glu Ile Leu Asn Asn Ile 850 855 860	2592
15	ATC TTA AAT TTA AGA TAT AAG GAT AAT AAT TTA ATA GAT TTA TCA GGA Ile Leu Asn Leu Arg Tyr Lys Asp Asn Asn Leu Ile Asp Leu Ser Gly 865 870 875	2640
	TAT GGG GCA AAG GTA GAG GTA TAT GAT GGA GTC GAG CTT AAT GAT AAA Tyr Gly Ala Lys Val Glu Val Tyr Asp Gly Val Glu Leu Asn Asp Lys 885 890 895	2688
20	AAT CAA TIT AAA TTA ACT AGT TCA GCA AAT AGT AAG ATT AGA GTG ACT Asn Gln Phe Lys Leu Thr Ser Ser Ala Asn Ser Lys Ile Arg Val Thr 900 910	2736
	CAA AAT CAG AAT ATC ATA TTT AAT AGT GTG TTC CTT GAT TTT AGC GTT Gln Asn Gln Asn Ile Ile Phe Asn Ser Val Phe Leu Asp Phe Ser Val 915	2784
25	AGC TTT TGG ATA AGA ATA CCT AAA TAT AAG AAT GAT GGT ATA CAA AAT Ser Phe Trp Ile Arg Ile Pro Lys Tyr Lys Asn Asp Gly Ile Gln Asn 930	2832
30	TAT ATT CAT AAT GAA TAT ACA ATA ATT AAT TGT ATG AAA AAT AAT TCG Tyr Ile His Asn Glu Tyr Thr Ile Ile Asn Cys Met Lys Asn Asn Ser 950 950 960	2880
	GGC TGG AAA ATA TCT ATT AGG GGT AAT AGG ATA ATA TGG ACT TTA ATT Gly Trp Lys Ile Ser Ile Arg Gly Asn Arg Ile Ile Trp Thr Leu Ile 965 970 975	2928
35	GAT ATA AAT GGA AAA ACC AAA TCG GTA TTT TTT GAA TAT AAC ATA AGA ASP Ile Asn Gly Lys Thr Lys Ser Val Phe Phe Glu Tyr Asn Ile Arg 980 990	2976
40	GAA GAT ATA TCA GAG TAT ATA AAT AGA TGG TIT TIT GTA ACT ATT ACT Glu Asp Ile Ser Glu Tyr Ile Asn Arg Trp Phe Phe Val Thr Ile Thr 1000 1005	3024
	AAT AAT TTG AAT AAC GCT AAA ATT TAT ATT AAT GGT AAG CTA GAA TCA Asn Asn Leu Asn Asn Ala Lys Ile Tyr Ile Asn Gly Lys Leu Glu Ser 1010 1015 1020	3072
45	AAT ACA GAT ATT AAA GAT ATA AGA GAA GTT ATT GCT AAT GGT GAA ATA Asn Thr Asp Ile Lys Asp Ile Arg Glu Val Ile Ala Asn Gly Glu Ile 1035 1040	3120
50	ATA TTT AAA TTA GAT GGT GAT ATA GAT AGA ACA CAA TTT ATT TGG ATG  Ile Phe Lys Leu Asp Gly Asp Ile Asp Arg Thr Gln Phe Ile Trp Met  1045  1055	3168
	AAA TAT TTC AGT ATT TTT AAT ACG GAA TTA AGT CAA TCA AAT ATT GAA Lys Tyr Phe Ser Ile Phe Asn Thr Glu Leu Ser Gln Ser Asn Ile Glu 1060 1065 1070	3216
55	GAA AGA TAT AAA ATT CAA TCA TAT AGC GAA TAT TTA AAA GAT TTT TGG Glu Arg Tyr Lys Ile Gln Ser Tyr Ser Glu Tyr Leu Lys Asp Phe Trp 1075 1080	3264

5		109	10				109	5	. 010	yı	c ly:	110	t Phe	e Asr	Ala	GGG Gly	3312
	110	5			-	111	0 7		Lys	wys	111	se: .5	Pro	Val	Gly	GAA Glu 1120	3360
10				CGT Arg	1129	5 -	- 4 -		9411	113	o O	Lys	Tyr	Ile	Asn 113	Tyr	3408
15				TAT Tyr 1140	•			<b>-</b>	1145	116	TIE	Arg	Arg	Lys 1150	Ser	Asn	3456
	TCT Ser	CAA Gln	TCT Ser 1155	ATA . Ile .	AAT Asn	GAT Asp	GAT Asp	ATA Ile 1160	GTT Val	AGA Arg	AAA Lys	GAA Glu	GAT Asp 1165	Tyr	ATA Ile	TAT Tyr	3504
20	CTA Leu	GA															3509

# (2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1169 amino acids (B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

55

25

30

35

40

45

											_	. •				•	p Asn 5
5										_	•				31	0	y Arg
														49:	5		o Glu
10													90	•			Gly
												/ 3					Asn 80
15											90				Lys	95	
															Glu 110		
20														125	Leu		
													740		Ile		Asn
25												T22			Leu		Ile 160
	Phe	Gly	P	ro (	31y	Pro 165	Val	Leu	Asn	Glu	Asn 170	Glu	Thr	Ile	Asp	11e 175	Gly

	Ile Gln Asn His Phe Ala Ser Arg Glu Gly Phe Gly Gly Ile Met Gln 180 185 190
5	Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Gln Glu 195 200 205
	Asn Lys Gly Ala Ser Ile Phe Asn Arg Arg Gly Tyr Phe Ser Asp Pro 210 215 220
10	Ala Leu Ile Leu Met His Glu Leu Ile His Val Leu His Gly Leu Tyr 235 240
	Gly Ile Lys Val Asp Asp Leu Pro Ile Val Pro Asn Glu Lys Lys Phe 245 250 255
15	Phe Met Gln Ser Thr Asp Ala Ile Gln Ala Glu Glu Leu Tyr Thr Phe 260 265 270
	Gly Gly Gln Asp Pro Ser Ile Ile Thr Pro Ser Thr Asp Lys Ser Ile 280 285
20	Tyr Asp Lys Val Leu Gln Asn Phe Arg Gly Ile Val Asp Arg Leu Asn 290 295 300
25	Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr 305 310 315 320  Lys Asn Lys Phe Lys Asp Lys Typ Lys Diversity 100 200
20	Lys Asn Lys Phe Lys Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly 325 330 335  Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu 340
30	Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lye Ile Ive
	Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys
35	Asn Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile
	Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile 405 410 415
40	Asn Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr 420 425 430
	Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp 435 440 445
<b>4</b> 5	Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser 450 460
	Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn 470 475 480  Tyr Ile Glu Asn Asp Dbe Dro Ile Asp Cl
50	Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp 485 490 495  Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr
	Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys
55	515 520 525

	Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln 530 540
5	Thr Phe Pro Leu Asp Ile Arg Asp Ile Ser Leu Thr Ser Ser Phe Asp 545 550 555 560
	Asp Ala Leu Leu Phe Ser Asn Lys Val Tyr Ser Phe Phe Ser Met Asp 565 570 575
10	Tyr Ile Lys Thr Ala Asn Lys Val Val Glu Ala Gly Leu Phe Ala Gly 580 585 590
	Trp Val Lys Gln Ile Val Asn Asp Phe Val Ile Glu Ala Asn Lys Ser 595 600 605
15	Asn Thr Met Asp Lys Ile Ala Asp Ile Ser Leu Ile Val Pro Tyr Ile 610 615 620
	Gly Leu Ala Leu Asn Val Gly Asn Glu Thr Ala Lys Gly Asn Phe Glu
20	Asn Ala Phe Glu Ile Ala Gly Ala Ser Ile Leu Leu Glu Phe Ile Pro 645 650 655
	Glu Leu Leu Ile Pro Val Val Gly Ala Phe Leu Leu Glu Ser Tyr Ile 660 665 670
25	Asp Asn Lys Asn Lys Ile Ile Lys Thr Ile Asp Asn Ala Leu Thr Lys 675 680 685
	Arg Asn Glu Lys Trp Ser Asp Met Tyr Gly Leu Ile Val Ala Gln Trp 690 695 700
30	Leu Ser Thr Val Asn Thr Gln Phe Tyr Thr Ile Lys Glu Gly Met Tyr 705 710 715 720
	Lys Ala Leu Asn Tyr Gln Ala Gln Ala Leu Glu Glu Ile Ile Lys Tyr 725 730 735
35	Arg Tyr Asn Ile Tyr Ser Glu Lys Glu Lys Ser Asn Ile Asn Ile Asp 740 745 750
	Phe Asn Asp Ile Asn Ser Lys Leu Asn Glu Gly Ile Asn Gln Ala Ile 755 760 765
40	Asp Asn Ile Asn Asn Phe Ile Asn Gly Cys Ser Val Ser Tyr Leu Met 770 780
	Lys Lys Met Ile Pro Leu Ala Val Glu Lys Leu Leu Asp Phe Asp Asn 790 795 800
45	Thr Leu Lys Lys Asn Leu Leu Asn Tyr Ile Asp Glu Asn Lys Leu Tyr 805 810 815
	Leu Ile Gly Ser Ala Glu Tyr Glu Lys Ser Lys Val Asn Lys Tyr Leu 820 825 830
50	Lys Thr Ile Met Pro Phe Asp Leu Ser Ile Tyr Thr Asn Asp Thr Ile 835 840 845
	Leu Ile Glu Met Phe Asn Lys Tyr Asn Ser Glu Ile Leu Asn Asn Ile 850 855
55	Ile Leu Asn Leu Arg Tyr Lys Asp Asn Asn Leu Ile Asp Leu Ser Gly 875 880

_	Tyr Gly Ala Lys Val Glu Val Tyr Asp Gly Val Glu Leu Asn Asp Lys 885 890 895
5	Asn Gln Phe Lys Leu Thr Ser Ser Ala Asn Ser Lys Ile Arg Val Thr 900 905 910
10	Gln Asn Gln Asn Ile Ile Phe Asn Ser Val Phe Leu Asp Phe Ser Val 915 920 925
10	Ser Phe Trp Ile Arg Ile Pro Lys Tyr Lys Asn Asp Gly Ile Gln Asn 930 935 940
	Tyr Ile His Asn Glu Tyr Thr Ile Ile Asn Cys Met Lys Asn Asn Ser 945 950 955 960
15	Gly Trp Lys Ile Ser Ile Arg Gly Asn Arg Ile Ile Trp Thr Leu Ile
	Asp Ile Asn Gly Lys Thr Lys Ser Val Phe Phe Glu Tyr Asn Ile Arg 980 985 990
20	Glu Asp Ile Ser Glu Tyr Ile Asn Arg Trp Phe Phe Val Thr Ile Thr 995 1000 1005
	Asn Asn Leu Asn Asn Ala Lys Ile Tyr Ile Asn Gly Lys Leu Glu Ser 1010 1015 1020
25	Asn Thr Asp Ile Lys Asp Ile Arg Glu Val Ile Ala Asn Gly Glu Ile 1025 1030 1035 1040
20	Ile Phe Lys Leu Asp Gly Asp Ile Asp Arg Thr Gln Phe Ile Trp Met 1045 1050 1055
30	Lys Tyr Phe Ser Ile Phe Asn Thr Glu Leu Ser Gln Ser Asn Ile Glu 1060 1065 1070
	Glu Arg Tyr Lys Ile Gln Ser Tyr Ser Glu Tyr Leu Lys Asp Phe Trp 1075 1080 1085
35	Gly Asn Pro Leu Met Tyr Asn Lys Glu Tyr Tyr Met Phe Asn Ala Gly 1090 1095 1100
10	Asn Lys Asn Ser Tyr Ile Lys Leu Lys Lys Asp Ser Pro Val Gly Glu 1105 1110 1115 1120
40	Ile Leu Thr Arg Ser Lys Tyr Asn Gln Asn Ser Lys Tyr Ile Asn Tyr 1125 1130 1135
45	Arg Asp Leu Tyr Ile Gly Glu Lys Phe Ile Ile Arg Arg Lys Ser Asn 1140 1145 1150
<b>4</b> 5	Ser Gln Ser Ile Asn Asp Asp Ile Val Arg Lys Glu Asp Tyr Ile Tyr 1155 1160 1165
	Leu
50	(2) INFORMATION FOR SEQ ID NO: 21:
	(i) SEQUENCE CHARACTERISTICS:
55	(A) LENGTH: 2574 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: double (D) TOPOLOGY: linear

	(ii) MOLECULE TYPE: DNA (genomic)
	(ix) FEATURE:
5	(A) NAME/KEY: CDS (B) LOCATION:12574
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:
10	
15	
20	
20	
25	
30	
35	
33	
40	
45	
50	
00	

5	ATG CCA GTT ACA ATA AAT AAT TTT AAT TAT AAT GAT CCT ATT GAT AAT Met Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile Asp Asn 1 5 10 15	4.8
	AAT AAT ATT ATG ATG GAG CCT CCA TTT GCG AGA GGT ACG GGG AGA Asn Asn Ile Ile Met Met Glu Pro Pro Phe Ala Arg Gly Thr Gly Arg 20 25 30	96
10	TAT TAT AAA GCT TTT AAA ATC ACA GAT CGT ATT TGG ATA ATA CCG GAA Tyr Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile Pro Glu 35 40 45	144
15	AGA TAT ACT TTT GGA TAT AAA CCT GAG GAT TTT AAT AAA AGT TCC GGT Arg Tyr Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser Gly 50 55	192
	ATT TTT AAT AGA GAT GTT TGT GAA TAT TAT	240
20	ACT AAT GAT AAA AAG AAT ATA TTT TTA CAA ACA ATG ATC AAG TTA TTT Thr Asn Asp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile Lys Leu Phe 85	288
25	AAT AGA ATC AAA TCA AAA CCA TTG GGT GAA AAG TTA TTA GAG ATG ATT Asn Arg Ile Lys Ser Lys Pro Leu Gly Glu Lys Leu Leu Glu Met Ile 105 110	336
	ATA AAT GGT ATA CCT TAT CTT GGA GAT AGA CGT GTT CCA CTC GAA GAG Ile Asn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro Leu Glu Glu 115 120 125	384
30	TTT AAC ACA AAC ATT GCT AGT GTA ACT GTT AAT AAA TTA ATC AGT AAT Phe Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile Ser Asn 130 135 140	432
35	CCA GGA GAA GTG GAG CGA AAA AAA GGT ATT TTC GCA AAT TTA ATA ATA Pro Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn Leu Ile Ile 150 150 160	480
	TTT GGA CCT GGG CCA GTT TTA AAT GAA AAT GAG ACT ATA GAT ATA GGT Phe Gly Pro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile Asp Ile Gly 165 170 175	S28
40	ATA CAA AAT CAT TIT GCA TCA AGG GAA GGC TTC GGG GGT ATA ATG CAA Ile Gln Asn His Phe Ala Ser Arg Glu Gly Phe Gly Gly Ile Met Gln 180 185	576
45	ATG AAG TTT TGC CCA GAA TAT GTA AGC GTA TTT AAT AAT GTT CAA GAA Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Gln Glu 195 200 205	624
70	AAC AAA GGC GCA AGT ATA TTT AAT AGA CGT GGA TAT TTT TCA GAT CCA Asn Lys Gly Ala Ser Ile Phe Asn Arg Arg Gly Tyr Phe Ser Asp Pro 215 220	672
50	GCC TTG ATA TTA ATG CAT GAA CTT ATA CAT GTT TTA CAT GGA TTA TAT Ala Leu Ile Leu Met His Glu Leu Ile His Val Leu His Gly Leu Tyr 235 240	720

	GGC ATT AAA GTA GAT GAT TTA CCA ATT GTA CCA AAT GAA AAA AAA TTT Gly Ile Lys Val Asp Asp Leu Pro Ile Val Pro Asn Glu Lys Lys Phe 245	768
5	TTT ATG CAA TCT ACA GAT GCT ATA CAG GCA GAA GAA CTA TAT ACA TTT Phe Met Gln Ser Thr Asp Ala Ile Gln Ala Glu Glu Leu Tyr Thr Phe 260 265 270	816
10	GGA GGA CAA GAT CCC AGC ATC ATA ACT CCT TCT ACG GAT AAA AGT ATC Gly Gly Gln Asp Pro Ser Ile Ile Thr Pro Ser Thr Asp Lys Ser Ile 275 280 285	864
45	TAT GAT AAA GTT TTG CAA AAT TTT AGA GGG ATA GTT GAT AGA CTT AAC Tyr Asp Lys Val Leu Gln Asn Phe Arg Gly Ile Val Asp Arg Leu Asn 290 295 300	912
15	AAG GTT TTA GTT TGC ATA TCA GAT CCT AAC ATT AAT ATT AAT ATA TAT Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr 305 310 320	960
20	AAA AAT AAA TTT AAA GAT AAA TAT AAA TTC GTT GAA GAT TCT GAG GGA Lys Asn Lys Phe Lys Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly 325 330 335	1008
	AAA TAT AGT ATA GAT GTA GAA AGT TTT GAT AAA TTA TAT AAA AGC TTA Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu 340 345 350	1056
25	ATG TTT GGT TTT ACA GAA ACT AAT ATA GCA GAA AAT TAT AAA ATA AAA Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys 355 360 365	1104
30	ACT AGA GCT TCT TAT TTT AGT GAT TCC TTA CCA CCA GTA AAA ATA AAA Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys 375 380	1152
	AAT TTA TTA GAT AAT GAA ATC TAT ACT ATA GAG GAA GGG TTT AAT ATA Asn Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile 395 390 395	1200
35	TCT GAT AAA GAT ATG GAA AAA GAA TAT AGA GGT CAG AAT AAA GCT ATA Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile 405 410 415	1248
40	AAT AAA CAA GCT TAT GAA GAA ATT AGC AAG GAG CAT TTG GCT GTA TAT Asn Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr 420 430	1296
	AAG ATA CAA ATG TGT AAA AGT GTT AAA GCT CCA GGA ATA TGT ATT GAT Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp 445	1344
45	GTT GAT AAT GAA GAT TTG TTC TTT ATA GCT GAT AAA AAT AGT TTT TCA Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser 450 450	1392
50	GAT GAT TTA TCT AAA AAC GAA AGA ATA GAA TAT AAT ACA CAG AGT AAT Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn 465 470 480	1440
	TAT ATA GAR AAT GAC TTC CCT ATA AAT GAA TTA ATT TTA GAT ACT GAT Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp 485 490 495	1488
55	TTA ATA AGT AAA ATA GAA TTA CCA AGT GAA AAT ACA GAA TCA CTT ACT Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr 500 505	1536

	GAT TTT AAT GTA GAT GTT CCA GTA TAT GAA AAA CAA CCC GCT ATA AAA Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys	1584
5	AAA ATT TTT ACA GAT GAA AAT ACC ATC TTT CAA TAT TTA TAC TCT CAG Lys lle Phe Thr Asp Glu Asn Thr lle Phe Gln Tyr Leu Tyr Ser Gln	1632
10	ACA TTT CCT CTA GAT ATA AGA GAT ATA AGT TTA ACA TCT TCA TTT GAT Thr Phe Pro Leu Asp Ile Arg Asp Ile Ser Leu Thr Ser Ser Phe Asp 545 550 550 560	1680
15	GAT GCA TTA TTA TTT TCT AAC AAA GTT TAT TCA TTT TTT TCT ATG GAT Asp Ala Leu Leu Phe Ser Asn Lys Val Tyr Ser Phe Phe Ser Met Asp 565 570 575	1728
	TAT ATT AAA ACT GCT AAT AAA GTG GTA GAA GCA GGA TTA TTT GCA GGT Tyr Ile Lys Thr Ala Asn Lys Val Val Glu Ala Gly Leu Phe Ala Gly 585 590	1776
20	TGG GTG AAA CAG ATA GTA AAT GAT TTT GTA ATC GAA GCT AAT AAA AGC Trp Val Lys Gin Ile Val Asn Asp Phe Val Ile Glu Ala Asn Lys Ser 595 600 605	1824
	AAT ACT ATG GAT AAA ATT GCA GAT ATA TCT CTA ATT GTT CCT TAT ATA Asn Thr Met Asp Lys Ile Ala Asp Ile Ser Leu Ile Val Pro Tyr Ile 610 620	1872
25	GGA TTA GCT TTA AAT GTA GGA AAT GAA ACA GCT AAA GGA AAT TTT GAA Gly Leu Ala Leu Asn Val Gly Asn Glu Thr Ala Lys Gly Asn Phe Glu 635 640	1920
30	AAT GCT TTT GAG ATT GCA GGA GCC AGT ATT CTA CTA GAA TTT ATA CCA Asn Ala Phe Glu Ile Ala Gly Ala Ser Ile Leu Leu Glu Phe Ile Pro 645 650 655	1968
	GAA CTT TTA ATA CCT GTA GTT GGA GCC TTT TTA TTA GAA TCA TAT ATT Glu Leu Leu Ile Pro Val Val Gly Ala Phe Leu Leu Glu Ser Tyr Ile 660 665 670	2016
35	GAC AAT AAA AAT AAA ATT ATT AAA ACA ATA GAT AAT GCT TTA ACT AAA Asp Asn Lys Asn Lys Ile Ile Lys Thr Ile Asp Asn Ala Leu Thr Lys 675 680 685	2064
40	AGA AAT GAA AAA TGG AGT GAT ATG TAC GGA TTA ATA GTA GCG CAA TGG Arg Asn Glu Lys Trp Ser Asp Met Tyr Gly Leu Ile Val Ala Gln Trp 690 695	2112
	CTC TCA ACA GTT AAT ACT CAA TTT TAT ACA ATA AAA GAG GGA ATG TAT Leu Ser Thr Val Asn Thr Gln Phe Tyr Thr Ile Lys Glu Gly Met Tyr 715 710 720	2160
45	AAG GCT TTA AAT TAT CAA GCA CAA GCA TTG GAA GAA ATA ATA AAA TAC Lys Ala Leu Asn Tyr Gln Ala Gln Ala Leu Glu Glu Ile Ile Lys Tyr 725 730 735	2208
50	AGA TAT AAT ATA TAT TCT GAA AAA GAA AAG TCA AAT ATT AAC ATC GAT Arg Tyr Asn Ile Tyr Ser Glu Lys Glu Lys Ser Asn Ile Asn Ile Asp 740	2256
30	TTT AAT GAT ATA AAT TCT AAA CTT AAT GAG GGT ATT AAC CAA GCT ATA Phe Asn Asp Ile Asn Ser Lys Leu Asn Glu Gly Ile Asn Gln Ala Ile 755 760 765	2304
55	GAT AAT ATA AAT TIT ATA AAT GGA TGT TCT GTA TCA TAT TTA ATG Asp Asn Ile Asn Asn Phe Ile Asn Gly Cys Ser Val Ser Tyr Leu Met 770 780	2352

5	785	-				790	72.0	447	GIU	Lys	195	Leu	Asp	Phe	Asp	AAT Asn 800	2400
			-,-	,-	805		200	ASII	TYL	810	Asp	Glu	Asn	Lys	Leu 815	_	2448
10			<b>01</b> 7	820	7.4	GIU	lyr	GIU	825	TCA Ser	Lys	Val	Asn	Eys 830	Tyr	Leu	2496
15	AAA Lys	ACC Thr	ATT Ile 835	ATG Met	CCG Pro	TTT Phe	wah	CTT Leu 840	TCA Ser	ATA Ile	TAT Tyr	ACC Thr	AAT Asn 845	GAT Asp	ACA Thr	ATA Ile	2544
	Leu	ATA Ile 850	GAA Glu	ATG Met	TTT Phe	Asn	AAA Lys 855	TAT Tyr	AAT Asn	AGC Ser							2574

- (2) INFORMATION FOR SEQ ID NO: 22:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 858 amino acids
    - (B) TYPE: amino acid

25

35

40

45

50

55

- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- 30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

	Me	t Pr 1	o Va	1 Th	r Ile	e As:	n Ası	n Phe	e Ası	n Ty:	r Ası	n As	p Pr	0 Ile	As 1	p Asn
5										,				30	G1;	y Arg
													4.5	5		Glu
10												60				Gly
											/5					Asn 80
15					Lys 85					90					95	
					Ser				103					110		
20					Pro								125			
					Ile		~~					140				
25											122					160
					Pro '					1,0				;	175	
30	Ile	Gln	Asn	His 180	Phe A	Ala :	Ser :	Arg (	31u   185	Gly	Phe (	Gly	Gly	Ile : 190	det.	Gln

	Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Gln Glu 195 200 205
5	Asn Lys Gly Ala Ser Ile Phe Asn Arg Arg Gly Tyr Phe Ser Asp Pro 210 215 220
	Ala Leu Ile Leu Met His Glu Leu Ile His Val Leu His Gly Leu Tyr 225 230 235 240
10	Gly Ile Lys Val Asp Asp Leu Pro Ile Val Pro Asn Glu Lys Lys Phe 245 250 255
	Phe Met Gln Ser Thr Asp Ala Ile Gln Ala Glu Glu Leu Tyr Thr Phe 260 265 270
15	Gly Gly Gln Asp Pro Ser Ile Ile Thr Pro Ser Thr Asp Lys Ser Ile 280 285
	Tyr Asp Lys Val Leu Gln Asn Phe Arg Gly Ile Val Asp Arg Leu Asn 290 295 300
20	Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr 305 310 315 320
	Lys Asn Lys Phe Lys Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly 325 330 335
25	Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu 340 345 350  Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys 355
	355 360 365  Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys 370 375
30	370 375 380  Asn Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile 385 390
35	Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Glv Gln Asp Lys Ala Tlo
30	Asn Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr
40	Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp
	Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser 450 460
45	Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn 465 470 475 480
	Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp 485 490 495
50	Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr 500 505 510
	Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys 515 520 525
55	Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln 530 540

	Thr Phe Pro Leu Asp Ile Arg Asp Ile Ser Leu Thr Ser Ser Phe Asp
5	Asp Ala Leu Leu Phe Ser Asn Lys Val Tyr Ser Phe Phe Ser Met Asp
	575
	Tyr Ile Lys Thr Ala Asn Lys Val Val Glu Ala Gly Leu Phe Ala Gly 585 590
10	Trp Val Lys Gln Ile Val Asn Asp Phe Val Ile Glu Ala Asn Lys Ser 595 600 605
	Asn Thr Met Asp Lys Ile Ala Asp Ile Ser Leu Ile Val Pro Tyr Ile 610 615 620
15	Gly Leu Ala Leu Asn Val Gly Asn Glu Thr Ala Lys Gly Asn Phe Glu 625 630 635 640
	Asn Ala Phe Glu Ile Ala Gly Ala Ser Ile Leu Leu Glu Phe Ile Pro 645 650 655
20	Glu Leu Leu Ile Pro Val Val Gly Ala Phe Leu Leu Glu Ser Tyr Ile 660 665 670
	Asp Asn Lys Asn Lys Ile Ile Lys Thr Ile Asp Asn Ala Leu Thr Lys 675 680 685
25	Arg Asn Glu Lys Trp Ser Asp Met Tyr Gly Leu Ile Val Ala Gln Trp 690 695 700
	Leu Ser Thr Val Asn Thr Gln Phe Tyr Thr Ile Lys Glu Gly Met Tyr 705 710 715 720
30	Lys Ala Leu Asn Tyr Gln Ala Gln Ala Leu Glu Glu Ile Ile Lys Tyr 725 730 735
	Arg Tyr Asn Ile Tyr Ser Glu Lys Glu Lys Ser Asn Ile Asn Ile Asp 740 745 750
35	Phe Asn Asp Ile Asn Ser Lys Leu Asn Glu Gly Ile Asn Gln Ala Ile 755 760 765
	Asp Asn Ile Asn Asn Phe Ile Asn Gly Cys Ser Val Ser Tyr Leu Met 770 780
40	Lys Lys Met Ile Pro Leu Ala Val Glu Lys Leu Leu Asp Phe Asp Asn 785 790 795 800
	Thr Leu Lys Lys Asn Leu Leu Asn Tyr Ile Asp Glu Asn Lys Leu Tyr 805 810 815
45	Leu Ile Gly Ser Ala Glu Tyr Glu Lys Ser Lys Val Asn Lys Tyr Leu 820 825 830
	Lys Thr Ile Met Pro Phe Asp Leu Ser Ile Tyr Thr Asn Asp Thr Ile 835 840 845
50	Leu Ile Glu Met Phe Asn Lys Tyr Asn Ser 850 855
	(2) INFORMATION FOR SECURING 22

# (2) INFORMATION FOR SEQ ID NO: 23:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1644 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: double

	Li 0 303 (
	(D) TOPOLOGY: linear
	(ii) MOLECULE TYPE: DNA (genomic)
5	(ix) FEATURE:
	(A) NAME/KEY: CDS (B) LOCATION:11644
10	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23;
15	
7.5	
20	
25	
20	
30	
35	
40	
45	
50	

_	1			5			C ASI	10	ASD A	sp Pro	o Ile .	GAT AA Asp As 15	n
5			2	0			25	FILE	ATA V	rg Gly	Thr (	GG AG Gly Ar	3
10		_	35		., .	40	)	Arg .	ite Ti	rp Ile 45	Ile P	CG GAI	1
		50		•		55	GIU	wah 1	e As	on Lys	Ser S	CC GGT er Gly	, ————————————————————————————————————
15	65		_	•	70	, 011	-7-	IYL A	75	O ASP	Tyr L	TAA AT nek us 08	240
20			•	AAG A Lys A 85			200	90	ur we	c ite	Lys Le	eu Phe 95	288
		_	100	TCA A Ser L			105	GIU L	ys Let	r ren	Glu Me 110	t Ile	336
25	ATA AA Ile As	11	5			120	vab 1	JIG W	ca var	125	Leu Gl	u Glu	384
30	TTT AA Phe As 13	0			135		****	di As	in Lys 140	Leu :	Ile Se	r Asn	432
	CCA GG Pro Gly 145	-		15	0	_,_ ,	Gay I	16 Ph	e Ala	Asn I	eu Ile	160	480
35	TTT GG/ Phe Gly			165			1	30 GI	u Thr	Ile A	sp Ile 175	Gly	528
40	ATA CAP		180			1	185	ry Pne	e GIA	Gly I	le Met 90	Gln	576
	ATG AAG Met Lys	195	•		,-	200	er ve	ar Phe	Asn	Asn Va 205	al Gln	Glu	624
<b>4</b> 5	AAC AAA Asn Lys 210	GGC Gly	GCA A Ala S	GT ATA	Phe 215	AAT A Asn A	GA CO	T GGA	TAT Tyr 220	TTT TO	CA GAT er Asp	CCA Pro	672

	GCC TTG ATA TTA ATG CAT GAA CTT ATA CAT GTT TTA CAT GGA TTA TAT Ala Leu Ile Leu Met His Glu Leu Ile His Val Leu His Gly Leu Tyr 230 235 240	720
5	GGC ATT AAA GTA GAT GAT TTA CCA ATT GTA CCA AAT GAA AAA AAA TTT Gly He Lys Val Asp Asp Leu Pro He Val Pro Asn Glu Lys Lys Phe 245 250 255	768
10	TTT ATG CAA TCT ACA GAT GCT ATA CAG GCA GAA GAA CTA TAT ACA TTT Phe Met Gln Ser Thr Asp Ala Ile Gln Ala Glu Glu Leu Tyr Thr Phe 260 265 270	816
45	GGA GGA CAA GAT CCC AGC ATC ATA ACT CCT TCT ACG GAT AAA AGT ATC Gly Gly Gln Asp Pro Ser Ile Ile Thr Pro Ser Thr Asp Lys Ser Ile 280 285	864
15	TAT GAT AAA GTT TTG CAA AAT TTT AGA GGG ATA GTT GAT AGA CTT AAC Tyr Asp Lys Val Leu Gln Asn Phe Arg Gly Ile Val Asp Arg Leu Asn 295 300	912
20	AAG GTT TTA GTT TGC ATA TCA GAT CCT AAC ATT AAT ATT AAT ATA TAT Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr 305 310 320	960
	AAA AAT AAA TTT AAA GAT AAA TAT AAA TTC GTT GAA GAT TCT GAG GGA Lys Asn Lys Phe Lys Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly 325 330 335	1008
25	AAA TAT AGT ATA GAT GTA GAA AGT TTT GAT AAA TTA TAT AAA AGC TTA Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu 340 345	1056
30	ATG TIT GGT TIT ACA GAA ACT AAT ATA GCA GAA AAT TAT AAA ATA AAA Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys 355 360 365	1104
	ACT AGA GCT TCT TAT TTT AGT GAT TCC TTA CCA CCA GTA AAA ATA AAA Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys 375 380	1152
35	AAT TTA TTA GAT AAT GAA ATC TAT ACT ATA GAG GAA GGG TTT AAT ATA Asn Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile 385 395 400	1200
40	TCT GAT AAA GAT ATG GAA AAA GAA TAT AGA GGT CAG AAT AAA GCT ATA Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile 405 410	1248
	AAT AAA CAA GCT TAT GAA GAA ATT AGC AAG GAG CAT TTG GCT GTA TAT Asn Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr 420 430	1296
45	AAG ATA CAA ATG TGT AAA AGT GTT AAA GCT CCA GGA ATA TGT ATT GAT Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp 435 440	1344
50	GTT GAT AAT GAA GAT TTG TTC TTT ATA GCT GAT AAA AAT AGT TTT TCA Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser 450 460	1392
	GAT GAT TTA TCT AAA AAC GAA AGA ATA GAA TAT AAT ACA CAG AGT AAT ASP ASP Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn 470 480	1440
55	TAT ATA GAA AAT GAC TTC CCT ATA AAT GAA TTA ATT TTA GAT ACT GAT Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp 485 490 495	1488

	TTA Leu	ATA Ile	A AGI Ser	Lys 500	ATA Ile	GAA Glu	TTA Leu	CCA Pro	AGT Ser 505	GAA Glu	AAT Asn	ACA Thr	GAA Glu	TCA Ser 510	CTT Leu	ACT Thr		1536
5			515	GTA Val	_			520	~ 7 ~	310	туз	GIN	525	Ala	Ile	Lys	:	1584
10		530		ACA Thr	GAT Asp	GAA Glu	AAT Asn 535	ACC Thr	ATC Ile	TTT Phe	CAA Gln	TAT Tyr 540	TTA Leu	TAC Tyr	TCT Ser	CAG Gln	1	1632
	ACA Thr 545	TTT	CCT Pro	CTA Leu													1	644
15	(2) IN	FORM	/ATIO	N FOF	R SEC	ID N	O: 24:											
	(i) SEQUENCE CHARACTERISTICS:																	
20	(A) LENGTH: 548 amino acids (B) TYPE: amino acid (D) TOPOLOGY: linear																	
25	(ii) MOLECULE TYPE: protein  (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:																	
	(xi	) SEC	UENC	CE DE	SCRI	PTION	l: SEC	N DI Q	O: 24	;								
30																		
35																		
40																		
40																		
45																		
50																		

				al Th	_	_				1	U				1	5
5				le Il	-				2.	,				3 (	)	_
				B Ala				•••					4.	5		
10				r Phe				•				60	)			
				n Arg							/ >					80
15	Thr	Ası	l Ası	b ra	Lys 85	Asn	Ile	Phe	Leu	Gln 90	Thr	Met	Ile	Lys	Leu 95	Phe
	Asn	Arg	Ile	Lys 100	Ser	Lys	Pro	Leu	Gly 105	Glu	Lys	Leu	Leu	Glu 110	Met	Ile
20	Ile	Asn	Gly 115	Ile	Pro	Tyr	Leu	Gly 120	Asp	Arg	Arg	Val	Pro 125	Leu	Glu	Glu
	Phe	Asn 130	Thr	Asn	Ile	Ala	Ser 135	Val	Thr	Val	Asn ·	Lys 140	Leu	Ile	Ser	Asn
25	Pro 145	Gly	Glu	Val	Glu	Arg 150	Lys	Lys	Gly	Ile	Phe 155	Ala	Asn	Leu	Ile	Ile 160
	Phe	Gly	Pro	Gly	Pro 165	Val	Leu	Asn	Glu	Asn 170	Glu	Thr	Ile		Ile 175	Gly
30	Ile	Gln	Asn	His 180	Phe .	Ala	Ser	Arg	Glu 185	Gly	Phe	Gly	Gly	Ile 190	Met	Gln
	Met	Lys	Phe 195	Cys	Pro (	Glu	Tyr	Val :	Ser	Val	Phe .	Asn	Asn 205	Val (	Gln	Glu

	Asn Lys Gly Ala Ser Ile Phe Asn Arg Arg Gly Tyr Phe Ser Asp Pro
5	Ala Leu Ile Leu Met His Glu Leu Ile His Val Leu His Gly Leu Tyr
	Gly Ile Lys Val Asp Asp Leu Pro Ile Val Pro Asn Glu Lys Lys Phe 245 250 255
10	Phe Met Gln Ser Thr Asp Ala Ile Gln Ala Glu Glu Leu Tyr Thr Phe 265 270
15	Gly Gln Asp Pro Ser Ile Ile Thr Pro Ser Thr Asp Lys Ser Ile 280 285
	Tyr Asp Lys Val Leu Gln Asn Phe Arg Gly Ile Val Asp Arg Leu Asn 290 295 300
20	Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr 310 315 320  Lys Asn Lys Phe Lys Asp Lys Tyr Lys Tyr Lys Asp L
	Lys Asn Lys Phe Lys Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly 325 330 335 Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu 340
25	Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tvr Lve Ile Luc
	Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys
30	Asn Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile
	Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile 405 410
35	Asn Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr 420 425 430
40	Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp 435 . 440 445  Val Asp Asp Gly Asp Ley Dbs Dbs 71
	Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser 450 455 460  Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn 470
45	Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp
	Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr
50	Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys
	Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln 530 535 540
55	Thr Phe Pro Leu 545

<sup>(2)</sup> INFORMATION FOR SEQ ID NO: 25:

	(i) SEQUENCE CHARACTERISTICS:	
5	<ul><li>(A) LENGTH: 2616 base pairs</li><li>(B) TYPE: nucleic acid</li><li>(C) STRANDEDNESS: double</li><li>(D) TOPOLOGY: linear</li></ul>	
	(ii) MOLECULE TYPE: DNA (genomic)	
10	(ix) FEATURE:	
	(A) NAME/KEY: CDS (B) LOCATION:12616	
15	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:	
20	ATG CAG TTC GTG AAC AAG CAG TTC AAC TAT AAG GAC CCT GTA AAC GGT Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly 10 15	48
	GTT GAC ATT GCC TAC ATC AAA ATT CCA AAC GCC GGC CAG ATG CAG CCG Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro 20 25 30	96
25	GTG AAG GCT TTC AAG ATT CAT AAC AAA ATC TGG GTT ATT CCG GAA CGC Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg 35 40 45	144
20	GAT ACA TTT ACG AAC CCG GAA GAA GGA GAC TTG AAC CCG CCG CCG GAA Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50	192
30	GCA AAG CAG GTG CCA GTT TCA TAC TAC GAT TCA ACC TAT CTG AGC ACA Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65	240
35	GAC AAC GAG AAG GAT AAC TAC CTG AAG GGA GTG ACC AAA TTA TTC GAG Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90 95	288
	CGT ATT TAT TCC ACT GAC CTG GGC CGT ATG CTG CTG ACC TCA ATC GTC Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val 100 105	336
40	CGC GGA ATC CCA TTT TGG GGT GGC AGT ACC ATT GAC ACG GAG TTG AAG Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys 115 120 125	384
<b>4</b> 5	GTT ATT GAC ACT AAC TGC ATT AAC GTG ATC CAA CCA GAC GGT AGC TAC Val lle Asp Thr Asn Cys lle Asn Val lle Gln Pro Asp Gly Ser Tyr 130 140	432
	AGA TCT GAA GAA CTT AAC CTC GTA ATC ATC GGG CCC TCC GCG GAC ATT Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160	480
50	ATC CAG TTT GAG TGC AAG AGC TTT GGC CAC GAA GTG TTG AAC CTG ACG Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175	528

CGT AAC GGT TAC GGC TCT ACT CAG TAC ATT CGT TTC AGC CCA GAC TTC Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180

	ACG TTC GGT TTC GAG GAG AGC CTG GAG GTT GAT ACC AAC CCG CTG TTG Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 200 205	624
5	GGT GCA GGC AAG TTC GCA ACT GAT CCA GCG GTG ACC CTG GCA CAC GAG Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 215 220	672
10	CTG ATC CAC GCC GGT CAT CGT CTG TAT GGC ATT GCG ATT AAC CCG AAC Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 230 230	720
	CGC GTG TTC AAG GTT AAC ACC AAC GCC TAC TAC GAG ATG AGT GGT TTA Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 255	768
15	GAA GTA AGC TTC GAG GAA CTG CGC ACG TTC GGT GGC CAT GAT GCG AAG Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270	816
20	TTT ATC GAC AGC TTG CAG GAG AAC GAG TTC CGT CTG TAC TAC TAC AAC Phe lle Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn 280 285	864
	AAG TTT AAA GAT ATT GCA AGT ACA CTG AAC AAG GCT AAG TCC ATT GTG Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300	912
25	GGT ACC ACT GCT TCA TTA CAG TAT ATG AAA AAT GTT TTT AAA GAG AAA Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 315 320	960
30	TAT CTC CTA TCT GAA GAT ACA TCT GGA AAA TTT TCG GTA GAT AAA TTA TYR Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 330 335	1008
	AAA TTT GAT AAG TTA TAC AAA ATG TTA ACA GAG ATT TAC ACA GAG GAT Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340 345	1056
35	AAT TTT GTT AAG TTT TTT AAA GTA CTT AAC AGA AAA ACA TAT TTG AAT Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355 360	1104
	TTT GAT AAA GCC GTA TTT AAG ATA AAT ATA GTA CCT AAG GTA AAT TAC Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 375 380	1152
40	ACA ATA TAT GAT GGA TIT AAT TTA AGA AAT ACA AAT TTA GCA GCA AAC Thr lle Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 385 390 400	1200
<b>4</b> 5	TTT AAT GGT CAA AAT ACA GAA ATT AAT AAT ATG AAT TTT ACT AAA CTA Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 410 415	1248
	AAA AAT TIT ACT GGA TIG TIT GAA TIT TAT AAG TIG CTA TGT GTA AGA Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 420 425	1296
50	435 440 Lys Gly Tyr Asn Lys	1344
55	GCA TTA AAT GAT TTA TGT ATC AAA GTT AAT AAT TGG GAC TTG TTT TTT Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe 450 460	1392

	AGT CCT TCA GAA GAT AAT TTT ACT AAT GAT CTA AAT AAA GGA GAA GAA Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu 475 480	1440
5	ATT ACA TCT GAT ACT AAT ATA GAA GCA GCA GAA GAA AAT ATT AGT TTA Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu 485 490 495	1488
10	GAT TTA ATA CAA CAA TAT TAT TTA ACC TTT AAT TTT GAT AAT GAA CCT Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro 500 510	1536
	GAA AAT ATT TCA ATA GAA AAT CTT TCA AGT GAC ATT ATA GGC CAA TTA Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu 515 520 525	1584
15	GAA CTT ATG CCT AAT ATA GAA AGA TTT CCT AAT GGA AAA AAG TAT GAG Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu 535 540	1632
20	TTA GAT AAA TAT ACT ATG TTC CAT TAT CTT CGT GCT CAA GAA TTT GAA Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu 550 555 560	1680
	CAT GGT AAA TCT AGG ATT GCT TTA ACA AAT TCT GTT AAC GAA GCA TTA His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu 565 570 575	1728
25	TTA AAT CCT AGT CGT GTT TAT ACA TTT TTT TCT TCA GAC TAT GTA AAG Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys 580 585 590	1776
	AAA GTT AAT AAA GCT ACG GAG GCA GCT ATG TTT TTA GGC TGG GTA GAA Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu 595 600 605	1824
30	CAA TTA GTA TAT GAT TTT ACC GAT GAA ACT AGC GAA GTA AGT ACT ACG Gin Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr 610 615 620	1872
35	GAT AAA ATT GCG GAT ATA ACT ATA ATT ATT CCA TAT ATA GGA CCT GCT Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala 630 640	1920
	TTA AAT ATA GGT AAT ATG TTA TAT AAA GAT GAT	1968
40	ATA TIT TCA GGA GCT GIT ATT CTG TTA GAA TIT ATA CCA GAG ATT GCA Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala 660 665 670	2016
45	ATA CCT GTA TTA GGT ACT TTT GCA CTT GTA TCA TAT ATT GCG AAT AAG Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys 675 680 685	2064
	GTT CTA ACC GTT CAA ACA ATA GAT AAT GCT TTA AGT AAA AGA AAT GAA Val Leu Thr Val Gin Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu 690 695 700	2112
50	AAA TGG GAT GAG GTC TAT AAA TAT ATA GTA ACA AAT TGG TTA GCA AAG Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys 705 710 720	2160
	GTT AAT ACA CAG ATT GAT CTA ATA AGA AAA AAA ATG AAA GAA GCT TTA Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu 725 730	2208

5				74	0			- 27	74	5	E 110	e Ası	а Туз	r Gl: 750	n Ty:	T AAT r Asn	2256
		_	75	5			-,	760	)	4 710	: ASI	ı Pne	765	Ile	? Ası	GAT Asp	2304
10		77	0				775	5		- ASI	LLYS	780	Met	Ile	Asn	ATA Ile	2352
15	785	, ,		-		790	4,7-		AGI	ser	795	reu	Met	Asn	Ser	800	2400
			•		805	AAA Lys	9	Deu	GIU	810	rne	Asp	Ala	Ser	Leu 815	Lys	2448
20	-			820	-,-	TAT Tyr		·yı	825	ASD	Arg	GIÀ	Thr	Leu 830	Ile	Gly	2496
25			835			AAA Lys	<b></b>	840	AGI	Asn	Asn	Thr	Leu 845	Ser	Thr	Asp	2544
		850					855	-7-	GTA Val	GAT . Asp .	Asn ·	CAA Gln 860	AGA :	TTA Leu	TTA Leu	TCT Ser	2592
30	ACA Thr 865	TTT Phe	ACT Thr	GAA Glu	Tyr	ATT I Ile I 870	AAG Lys	TAA									2616

- (2) INFORMATION FOR SEQ ID NO: 26:
- 35 (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 872 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

40

45

50

55

- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

	Met 1	Gln	Phe	Val	Asn 5	Lys	Gln	Phe	Asn	Tyr 10	Lys	yab	Pro	Val	Asn 15	Gly
5	Val	qeA	Ile	Ala 20	Tyr	Ile	Lys	Ile	Pro 25	Asn	Ala	Gly	Gln	Met 30	Gln	Pro
	Val	Lys	Ala 35	Phe	Lys	Ile	His	Asn 40	Lys	Ile	Trp	Val	Ile 45	Pro	Glu	Arg
10	Asp	Thr 50	Phe	Thr	Asn	Pro	Glu 55	Glu	Gly	Ąsp	Leu	Asn 60	Pro	Pro	Pro	G1u
	Ala 65	Lys	Gln	Val	Pro	Val 70	Ser	Tyr	Tyr	Asp	Ser 75	Thr	Tyr	Leu	Ser	Thr 80
15	Asp	Asn	Glu	Lys	Asp 85	Asn	Tyr	Leu	Lys	90 90	Val	Thr	Lys	Leu	Phe 95	Glu
	Arg	Ile	Tyr	<i>Ser</i> 100	Thr	Asp	Leu	Gly	Arg 105	Met	Leu	Leu	Thr	<i>Ser</i> 110	Ile	Val

	Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys 115 120 125
5	Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 135 140
	Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160
10	Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175
	Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190
15	Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205
	Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 215 220
20	Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 235 240
	Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255
25	Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270
	Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn 275 280 285
30	Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300
	Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 305 310 320
35	Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 330 335
	Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340 345 350
40	Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355 360 365
	Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 375 380
45	Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Asn 395 390 395 400
	Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 415
50	Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 420 425 430 Gly Ile Ile Thr Ser Lys Thr Lys Ser Lys Thr Lys
	Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys 435 440 445
55	Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe 450 460

	Ser Pro S 465	er Glu As	p Asn Phe 470	Thr Asn Asp	Leu Asn L	ys Gly Glu Glu 480
5				490		sn Ile Ser Leu 495
				303		sp Asn Glu Pro 510
10		_		320	53	le Gly Gln Leu 25
					540	s Lys Tyr Glu
15					222	n Glu Phe Glu 560
				370		n Glu Ala Leu 575
20		<del>-</del>		262		p Tyr Val Lys 590 Y Trp Val Glu
25			`	,,,,	50	y Trp Val Glu 5 1 Ser Thr Thr
20			Ile Thr I		620	Gly Pro Ala
30				•	35	640 Gly Ala Leu
	Ile Phe Ser	Gly Ala	Val Ile L		he Ile Pro	655 Glu Ile Ala 670
35	Ile Pro Val 675	Leu Gly	Thr Phe Al	la Leu Val S BO	er Tyr Ile 685	Ala Asn Lys
			933		700	Arg Asn Glu
40	Lys Trp Asp	•	20	71	1.5	720
	Val Asn Thr			730		735
45	Glu Asn Gln . Gln Tyr Thr			/43		750
	755 Leu Ser Ser		,,	U	765	
50	Asn Lys Phe 1		.,,		780	
	785 Ile Pro Tyr (	Sly Val L		79	<b>5</b>	800
55		805		810	-	815

	<b>3</b>		_													
							Ile		023					830		
5							Asp						845			
	Ile	Pro 850	Phe	Gln	Leu	Ser	Lys 855	Tyr	Val	Asp	Asn	Gln 860	Arg	Leu	Leu	Ser
10	Thr 865	Phe	Thr	Glu	Tyr	11e 870	Lys	*								
	(2) IN	FORM	1ATIO	N FOF	R SEQ	ID N	O: 27:									
15	(i)	SEQ	UENC	E CH	ARAC	TERIS	STICS:	;								
20		(B) (C)	TYPE STRA	: nucle NDED	574 ba eic acid NESS /: linea	d S: doul										
	(ii)	MOL	ECUL	E TYF	E: DN	IA (ge	nomic	)								
25							: SEQ		D: <b>27</b> :							
30																
25																
35																
40																

	ATGCCGGTTA CCATCAACAA CTTCAACTAC AACGACCCGA TCGACAACAA CAACATCATC	60
	ATGATGGAAC CGCCGTTCGC ACGTGGTACC GGTCGTTACT ACAAGGCTTT CAAGATCACC	
5	GACCGTATCT GGATCATCCC GGAACGTTAC ACCTTCGGTT ACAAACCTGA GGACTTCAAC	120
	AAGAGTAGCG GGATTTTCAA TCGTGACGTC TGCGAGTACT ATGATCCAGA TTATCTGAAT	180
	ACCAACGATA AGAAGAACAT ATTCCTTCAG ACTATGATCA AGTTATTTAA TAGAATCAAA	240
10	TCAAAACCAT TGGGTGAAAA GTTATTAGAG ATGATTATAA ATGGTATACC TTATCTTGGA	300
	GATAGACGTG TTCCACTCGA AGAGTTTAAC ACAAACATTG CTAGTGTAAC TGTTAATAAA	360
•	TTAATCAGTA ATCCAGGAGA AGTGGAGCGA AAAAAAGGTA TTTTCGCAAA TTTAATAATA	420
15	TTTGGACCTG GGCCAGTTTT AAATGAAAAT GAGACTATAG ATATAGGTAT ACAAAATCAT	480
	TTTGCATCAA GGGAAGGCTT CCCCCCTATA AGGACTATAG ATATAGGTAT ACAAAATCAT	540
	TTTGCATCAA GGGAAGGCTT CGGGGGTATA ATGCAAATGA AGTTTTGCCC AGAATATGTA	600
20	AGCGTATTTA ATAATGTTCA AGAAAACAAA GGCGCAAGTA TATTTAATAG ACGTGGATAT	660
	TTTTCAGATC CAGCCTTGAT ATTAATGCAT GAACTTATAC ATGTTTTACA TGGATTATAT	720
	GGCATTAAAG TAGATGATTT ACCAATTGTA CCAAATGAAA AAAAATTTTT TATGCAATCT	780
	ACAGATGCTA TACAGGCAGA AGAACTATAT ACATTTGGAG GACAAGATCC CAGCATCATA	840
25	ACTCCTTCTA CGGATAAAAG TATCTATGAT AAAGTTTTGC AAAATTTTAG AGGGATAGTT	900
	GATAGACTTA ACAAGGTTTT AGTTTGCATA TCAGATCCTA ACATTAATAT TAATATATAT	960
	AAAAATAAAT TTAAAGATAA ATATAAATTC GTTGAAGATT CTGAGGGAAA ATATAGTATA	1020
30	GATGTAGAAA GTTTTGATAA ATTATATAAA AGCTTAATGT TTGGTTTTAC AGAAACTAAT	1080
	ATAGCAGAAA ATTATAAAAT AAAAACTAGA GCTTCTTATT TTAGTGATTC CTTACCACCA	1140
	GTAAAAATAA AAAATTTATT AGATAATGAA ATCTATACTA TAGAGGAAGG GTTTAATATA	1200
35	- <del></del>	

	TCTGATAAAG ATATGGAAAA AGAATATAGA GGTCAGAATA AAGCTATAAA TAAACAAGCT	
	TATALAGA GGICAGAATA AAGCTATAAA TAAACAAGCT	1260
5	TATGAAGAA TTAGCAAGGA GCATTTGGCT GTATATAAGA TACAAATGTG TAAAAGTGTT	1320
	AAAGCTCCAG GAATATGTAT TGATGTTGAT AATGAAGATT TGTTCTTTAT AGCTGATAAA	1380
	AATAGTTTTT CAGATGATTT ATCTAAAAAC GAAAGAATAG AATATAATAC ACAGAGTAAT	
	TATATAGAAA ATGACTTCCC TATAAATGAA TTAATTTTAG ATACTGATTT AATAAGTAAA	1440
10		1500
	ATAGAATTAC CAAGTGAAAA TACAGAATCA CTTACTGATT TTAATGTAGA TGTTCCAGTA	1560
	TATGAAAAC AACCCGCTAT AAAAAAATT TITACAGATG AAAATACCAT CTTTCAATAT	1620
15	TTATACTCTC AGACATTTCC TCTAGATATA AGAGATATAA GTTTAACATC TTCATTTGAT	1680
15	GATGCATTAT TATTTTCTAA CAAAGTTTAT TCATTTTTTT CTATGGATTA TATTAAAACT	
	GCTAATAAAG TGGTAGAAGC AGGATTATTT GCAGGTTGGG TGAAACAGAT AGTAAATGAT	1740
	TTTGTAATCG AAGCTAATAA AAGCAATACT ATGGATAAAA TTGCAGATAT ATCTCTAATT	1800
20		1860
	GTTCCTTATA TAGGATTAGC TTTAAATGTA GGAAATGAAA CAGCTAAAGG AAATTTTGAA	1920
	AATGCTTTTG AGATTGCAGG AGCCAGTATT CTACTAGAAT TTATACCAGA ACTTTTAATA	1980
	CCTGTAGTTG GAGCCTTTTT ATTAGAATCA TATATTGACA ATAAAAATAA AATTATTAAA	2040
25	ACAATAGATA ATGCTTTAAC TAAAAGAAAT GAAAAATGGA GTGATATGTA CGGATTAATA	2100
	GTAGCGCAAT GGCTCTCAAC AGTTAATACT CAATTTTATA CAATAAAAGA GGGAATGTAT	
	AAGGCTTTAA ATTATCAAGC ACAAGCATTG GAAGAAATAA TAAAATACAG ATATAATATA	2160
30		2220
	TATTCTGARA ARGARARGTC ARRTATTRAC ATCGRTTTTA ATGRTRATARA TTCTARACTT	2280
	AATGAGGGTA TTAACCAAGC TATAGATAAT ATAAATAATT TTATAAATGG ATGTTCTGTA	2340
	TCATATTTAA TGAAAAAAT GATTCCATTA GCTGTAGAAA AATTACTAGA CTTTGATAAT	2400
35	ACTCTCAAAA AAAATTTGTT AAATTATATA GATGAAAATA AATTATATTT GATTGGAAGT	2460
	GCAGAATATG AAAAATCAAA AGTAAATAAA TACTTGAAAA CCATTATGCC GTTTGATCTT	
	TCAATATATA CCAATGATAC AATACTAATA GAAATGTTTA ATAAATATAA TAGC	2520
	THE TABLE OF THE TABLE TAGE	2574

40

45

35

- (2) INFORMATION FOR SEQ ID NO: 28:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 2574 base pairs (B) TYPE: nucleic acid

    - (C) STRANDEDNESS: double
    - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

55

	ATGCCAGTTA	CAATAAATAA	TTTTAATTAT	AATGATCCTA	TTGATAATAA	TAATATTATT	60
	ATGATGGAGC	CTCCATTTGC	GAGAGGTACG	GGGAGATATT	ATAAAGCTTT	TAAAATCACA	120
5	GATCGTATTT	GGATAATACC	GGAAAGATAT	ACTTTTGGAT	ATAAACCTGA	GGATTTTAAT	180
	AAAAGTTCCG	GTATTTTTAA	TAGAGATGTT	TGTGAATATT	ATGATCCAGA	TTACTTAAAT	240
10							
15							
20							
25							
30							
35							
40							
45							
50							
50							

	ACTAATGATA AAAAGAATAT ATTITTACAA ACAATGATCA AGTTATITAA TAGAATCAAA	300
5	TCAAAACCAT TGGGTGAAAA GTTATTAGAG ATGATTATAA ATGGTATACC TTATCTTGGA	
	GATAGACGTG TTCCACTCGA AGAGTTTAAC ACAAACATTG CTAGTGTAAC TGTTAATAAA	
	TTAATCAGTA ATCCAGGAGA AGTGGAGCGA AAAAAAGGTA TTTTCGCAAA TTTAATAATA	
10	TTTGGACCTG GGCCAGTTTT AAATGAAAAT GAGACTATAG ATATAGGTAT ACAAAATCAT	480
	TTTGCATCAA GGGAAGGCTT CGGGGGTATA ATGCAAATGA AGTTTTGCCC AGAATATGTA	
	AGCGTATTTA ATAATGTTCA AGAAAACAAA GGCGCAAGTA TATTTAATAG ACGTGGATAT	600
	TTTTCAGATC CAGCCTTGAT ATTAATGCAT GAACTCATCC ACGTCCTCCA CGGTCTCTAC	
15	GGTATCAAAG TAGACGACCT CCCGATCGTC CCGAACGAAA AAAAATTCTT CATGCAGAGC	
	ACCGACGCAA TCCAGGCAGA AGAACTCTAC ACCTTCGGTG GTCAGGACCC GAGCATCATC	780
20	ACCCCGAGCA CCGACAAAAG CATCTACGAC AAAGTCCTCC AGAACTTCCG TGGTATCGTC	840
	GACCGTCTCA ACAAAGTCCT CGTCTGCATC AGCGACCCGA ACATCAACAT CAACATCTAC	900
	AAAAACAAAT TCAAAGACAA ATACAAATTC GTCGAAGACA GCGAAGGTAA ATACAGCATC	960
25	GACGTCGAGA GCTTCGACAA ACTCTACAAA AGCCTCATGT TCGGTTTCAC CGAAACCAAC	1020
	ATCGCAGAAA ACTACAAAAT CAAAACCCGT GCAAGCTACT TCAGCGACAG CCTCCCGCCG	1080
	GTCAAAATCA AAAACCTCCT CGACAACGAA ATCTACACCA TCGAAGAAGG TTTCAACATC	1140
30	AGCGACAAAG ACATGGAAAA AGAATACCGT GGTCAGAACA AAGCAATCAA CAAACAAGCT	1200
	TACGAAGAAA TCAGCAAAGA ACACCTCGCA GTCTACAAAA TCCAGATGTG CAAAAGCGTC	1260
	AAAGCACCGG GTATCTGCAT CGACGTTGAC AACGAAGACC TCTTCTTCAT CGCAGACAAA	1320
	AACAGCTTCA GCGACGACCT CAGCAAAAAC GAACGTATCG AATACAACAC CCAGAGCAAC	1380 1440
35	TACATCGAAA ACGACTTCCC GATCAACGAA CTCATCCTCG ACACCGACCT CATCAGCAAA	1500
	ATCGAACTCC CGAGCGAAAA CACCGAAAGC CTCACCGACT TCAACGTTGA CGTCCCGGTC	1560
	TACGAAAAAC AGCCGGCAAT CAAAAAAATC TTCACCGACG AAAACACCAT CTTCCAGTAC	1620
40	CTCTACAGCC AGACCTTCCC GCTAGATATA AGAGATATAA GTTTAACATC TTCATTTGAT	1680
	GATGCATTAT TATTTTCTAA CAAAGTTTAT TCATTTTTTT CTATGGATTA TATTAAAACT	1740
	GCTAATAAAG TGGTAGAAGC AGGATTATTT GCAGGTTGGG TGAAACAGAT AGTAAATGAT	1800
45	TTTGTAATCG AAGCTAATAA AAGCAATACT ATGGATAAAA TTGCAGATAT ATCTCTAATT	1860
45	GTTCCTTATA TAGGATTAGC TTTAAATGTA GGAAATGAAA CAGCTAAAGG AAATTTTGAA	1920
50	AATGCTTTTG AGATTGCAGG AGCCAGTATT CTACTAGAAT TTATACCAGA ACTTTTAATA	1980
	CCTGTAGTTG GAGCCTTTTT ATTAGAATCA TATATTGACA ATAAAAATAA AATTATTAAA	2040
	ACAATAGATA ATGCTTTAAC TAAAAGAAAT GAAAAATGGA GTGATATGTA CGGATTAATA	2100
	GTAGCGCAAT GGCTCTCAAC AGTTAATACT CAATTTTATA CAATAAAAGA GGGAATGTAT	2160
55	AAGGCTTTAA ATTATCAAGC ACAAGCATTG GAAGAAATAA TAAAATACAG ATATAATATA	2220
	TATTCTGAAA AAGAAAAGTC AAATATTAAC ATCGATTTTA ATGATATAAA TTCTAAACTT	2280

	aatgagggta	TTAACCAAGC	TATAGATAAT	ATAAATAATT	TTATAAATGG	ATGTTCTGTA	2340
•	<b>TCATATTTAA</b>	TGAAAAAAAT	GATTCCATTA	GCTGTAGAAA	AATTACTAGA	CTTTGATAAT	2400
,	ACTOTORARA	AAAATTTGTT	ATATTATAA	GATGAAAATA	AATTATATTT	GATTGGAAGT	2460
C	CAGAATATG	AAAAATCAAA	AGTAAATAAA	TACTTGAAAA	CCATTATGCC	GTTTGATCTT	2520
7	CANTATATA	CCAATGATAC	AATACTAATA	GAAATGTTTA	ATAAATATAA	TAGC	2574

Claims

10

15

20

25

40

45

50

1. A polypeptide composing first and second domains, wherein

said first domain is a clostridial neurotoxin light chain or a fragment or variant of said clostridial neurotoxin light chain, wherein said first domain is capable of cleaving one or more vesicle or plasma-membrane associated proteins essential to exocytosis,

said second domain is a clostridial neurotoxin heavy chain H<sub>N</sub> portion or a fragment or variant of said clostridial neurotoxin heavy chain H<sub>N</sub> portion, wherein said second domain is capable of (i) translocating the polypeptide into a cell or (ii) increasing the solubility of the polypeptide compared to the solubility of the first domain on its own or (iii) both translocating the polypeptide into a cell and increasing the solubility of the polypeptide compared to the solubility of the first domain on its own,

wherein said second domain lacks an intact portion designated  $H_{C}$  of a clostridial neurotoxin heavy chain, and said polypeptide is a single chain polypeptide.

- 2. A polypeptide according to Claim 1 wherein the clostridial toxin is a botulinum toxin. 30
  - 3. A polypeptide according to Claim 1 or Claim 2 wherein the first domain exhibits endopeptidase activity specific for a substrate selected from one or more of SNAP-25, synaptobrevin/VAMP and syntaxin.
- 4. A polypeptide according to Claim 1 wherein the clostridial toxin heavy chain H<sub>N</sub> portion is from a botulinum toxin. 35
  - 5. A polypeptide according to any of Claims 1-4 further comprising a third domain that binds of the polypeptide to a cell, by binding of the third domain directly to a cell or by binding of the third domain to a ligand or to ligands that
  - 6. A polypeptide according to Claim 5 wherein said third domain is for binding the polypeptide to an immunoglobulin.
  - 7. A polypeptide according to Claim 6 wherein said third domain is a tandem repeat synthetic IgG binding domain derived from domain  $\beta$  of Staphylococcal protein A.
  - 8. A polypeptide according to Claim 5 wherein said third domain comprises an amino acid sequence that binds to a cell surface receptor.
  - 9. A polypeptide according to Claim 8 wherein said third domain is insulin-like growth factor-1 (IGF-1).
  - 10. A polypeptide according to any preceding claim comprising a botulinum toxin light chain or a fragment or a variant of a botulinum toxin light chain and a portion designated  $H_{N}$  of a botulinum toxin heavy chain.
- 11. A polypeptide according to Claim 10 wherein one or both of (a) the toxin light chain or fragment or variant of toxin light chain and (b) the portion of the toxin heavy chain are of botulinum toxin type A. 55
  - 12. A polypeptide according to Claim 11 wherein the botulinum toxin type A light chain variant has at residue 2 a glutamate, at residue 26 a lysine, and at residue 27 a tyrosine.

- 13. A polypeptide according to Claim 10 wherein one or both of (a) the toxin light chain or fragment or variant of toxin light chain and (b) the portion of the toxin heavy chain are of botulinum toxin type B.
- 14. A polypeptide according to any of Claims 1-9 comprising a botulinum toxin light chain or a fragment or a variant of a botulinum toxin light chain and at least 100 N-terminal amino acids of a botulinum toxin heavy chain.

5

15

25

30

40

- 15. A polypeptide according to Claim 14 comprising a botulinum toxin type B light chain, or a fragment or variant thereof, and 107 N-terminal amino acids of a botulinum toxin type B heavy chain.
- 16. A polypeptide according to Claim 11 or 12 comprising at least 423 of the N-terminal amino acids of botulinum toxin type A heavy chain.
  - 17. A polypeptide according to Claim 16 comprising a botulinum toxin type A light chain and 423 N-terminal amino acids of a botulinum toxin type A heavy chain.
  - 18. A polypeptide according to Claim 16 comprising a botulinum toxin type A light chain variant wherein residue 2 is a glutamate, residue 26 is a lysine and residue 27 is a tyrosine, and 423 N-terminal amino acids of a botulinum toxin type A heavy chain.
- 20 19. A polypeptide according to Claim 13 comprising at least 417 of the N-terminal amino acids of botulinum toxin type B heavy chain.
  - 20. A polypeptide according to Claim 19 comprising a botulinum toxin type B light chain and 417 N-terminal amino acids of a botulinum toxin type B heavy chain.
  - 21. A polypeptide according to any of Claims 10-20 lacking a portion designated H<sub>C</sub> of a botulinum toxin heavy chain.
  - 22. A polypeptide according to any of Claims 1-4, wherein the second domain is not capable of binding to cell surface receptors.
  - 23. A polypeptide according to any preceding claim comprising a variant of a clostridial toxin and further comprising a site for cleavage by a proteolytic enzyme, which cleavage site is not present in the native toxin.
- 24. A polypeptide according to Claim 23 comprising a variant of a clostridial toxin light chain and further comprising a site for cleavage by a proteolytic enzyme, which cleavage site is not present in the native toxin light chain.
  - 25. A polypeptide according to Claim 23 or 24 comprising a variant of a clostridial toxin heavy chain H<sub>N</sub> portion and further comprising a site for cleavage by a proteolytic enzyme, which cleavage site is not present in the native toxin heavy chain H<sub>N</sub> portion.
  - 26. A polypeptide according to Claim 23, 24 or 25 obtainable by modification of a DNA encoding the polypeptide so as to introduce one or more nucleotides coding for the cleavage site.
- 27. A fusion protein comprising a fusion of (a) a polypeptide according to any of Claims 1-26 with (b) a second polypeptide being a polypeptide or oligopeptide that binds to an affinity matrix so as to enable purification of the fusion protein using said matrix.
  - 28. A fusion protein according to Claim 27 wherein said second polypeptide binds to a chromatography column.
- 29. A fusion protein according to Claim 28 wherein said second chromatography column is an affinity matrix of glutathione sepharose.
  - 30. A fusion protein according to Claim 27, 28 or 29 wherein a specific protease cleavage site is incorporated between the first and second polypeptides, said protease site enabling proteolytic separation of first and second polypeptides.
  - 31. A composition comprising a polypeptide according to any preceding claim, said composition being non-toxic in vivo.

- 32. A composition according to Claim 31 or a polypeptide according to any of Claims 1-26 or a fusion protein according to Claim 27, 28, 29 or 30 for use as a positive control in a toxin assay.
- 33. A composition according to Claim.31 or a polypeptide according to any of Claims 1-26 or a fusion protein according to Claim 27, 28, 29 or 30 for use as a vaccine against clostridial toxin.
- 34. A composition according to Claim 31 or a polypeptide according to any of Claims 1-26 or a fusion protein according to Claim 27, 28, 29, or 30 for in vivo use.
- 35. A pharmaceutical composition comprising a composition according to Claim 31, a polypeptide according to any of claims 1-26 or a fusion protein according to Claim 27, 28, 29 or 30, in combination with a pharmaceutically acceptable carrier.
  - 36. A nucleic acid encoding a polypeptide or a fusion protein according to any of Claims 1-30.
  - 37. A nucleic acid encoding a polypeptide or a fusion protein according to Claim 36 and comprising nucleotides encoding residues 1-448 of a botulinum toxin type A light chain.
- 38. A nucleic acid according to Claim 36 or 37 comprising nucleotides encoding residues 1-423 of a botulinum toxin type A heavy chain H<sub>N</sub> domain.
  - 39. A nucleic acid encoding a polypeptide or a fusion protein according to Claim 36 and comprising nucleotides encoding residues 1-470 of a botulinum toxin type B light chain.
- 40. A nucleic acid encoding a polypeptide or a fusion protein according to Claim 36 or 39 comprising nucleotides encoding residues 1-417 of a botulinum toxin type B heavy chain H<sub>N</sub> domain.
  - **41.** A nucleic acid according to any of Claims 36-40 comprising nucleotides encoding a restriction endonuclease cleavage site not present in native clostridial toxin sequence.
  - **42.** A nucleic acid according to Claim 41 obtainable by modification of a nucleotide encoding a polypeptide or fusion protein according to any of claims 1-30 so as to introduce said cleavage site.
  - 43. A DNA according to any of Claims 36-42.
  - 44. A DNA selected from SEQ ID No:s 1, 3, 5, 9, 11, 13, 15, 17, 19, 21, 27 and 28.
  - **45.** A method of manufacture of a polypeptide according to any of Claims 1-26 comprising expressing in a host cell a nucleic acid according to any of Claims 36-44 and recovering the polypeptide.
  - **46.** A method of manufacture of a polypeptide according to any of Claims 1-26 comprising expressing in a host cell a nucleic acid encoding a fusion protein according to Claim 27, 28, 29 or 30, purifying the fusion protein by eluting the fusion protein through an affinity matrix adapted to retain the fusion protein and eluting through said matrix a ligand adapted to displace the fusion protein, and recovering the fusion protein.
  - 47. A method of manufacture according to Claims 45 or 46 in which the nucleic acid is DNA.
  - 48. A cell expressing a polypeptide or fusion protein according to any of Claims 1-30.

### Patentansprüche

5

15

30

35

40

45

50

- Polypeptid, umfassend erste und zweite Domänen,
  - worin die erste Domäne die leichte Kette eines Clostridienneurotoxins oder ein Fragment oder eine Variante der leichten Kette des Clostridienneurotoxins ist, worin die erste Domäne einen oder mehrere Vesikel- oder Plasmamembranassoziierte Proteine, die für die Exocytose wesentlich sind, spalten kann,
  - worin die zweite Domäne ein  $H_N$ -Teil der schweren Kette eines Clostridienneurotoxins oder ein Fragment oder eine Variante des  $H_N$ -Teils der schweren Kette des Clostridienneurotoxins ist, worin die zweite Domäne (i)

das Polypeptid in eine Zelle translozieren kann oder (ii) die Löslichkeit des Polypeptids im Vergleich zur Löslichkeit der ersten Domäne alleine erhöhen kann oder (iii) sowohl das Polypeptid in eine Zelle translozieren als auch die Löslichkeit des Polypeptids im Vergleich zur Löslichkeit der ersten Domäne alleine erhöhen kann,

worin der zweiten Domäne ein zweiter, als  $H_{\mathbb{C}}$ -bezeichneter intakter Teil der schweren Kette des Clostridienneurotoxins fehlt, und

worin das Polypeptid ein Einzelkettenpolypeptid ist.

5

20

25

35

50

55

- 2. Polypeptid nach Anspruch 1, worin das Clostridientoxin ein Botulinustoxin ist.
- Polypeptid nach Anspruch 1 oder Anspruch 2, worin die erste Domäne eine Endopeptidaseaktivität aufweist, die spezifisch für ein Substrat ist, ausgewählt unter einem oder mehreren von SNAP-25, Synaptobrevin/VAMP und Syntaxin
- Polypeptid nach Anspruch 1, worin der H<sub>N</sub>-Teil der schweren Kette des Clostridienneurotoxins von einem Botulinustoxin ist.
  - 5. Polypeptid nach einem der Ansprüche 1 bis 4, zusätzlich umfassend eine dritte Domäne, die von dem Polypeptid an eine Zelle bindet, und zwar indem die dritte Domäne direkt an eine Zelle bindet oder indem die dritte Domäne an einen Ligand oder an Liganden bindet, der bzw. die an eine Zelle bindet bzw. binden.
  - 6. Polypeptid nach Anspruch 5, worin die dritte Domäne für die Bindung des Polypeptids an ein Immunglobulin ist.
  - Polypeptid nach Anspruch 6, worin die dritte Domäne eine synthetische Tandemrepeat- IgG-Bindungsdomäne ist, abgeleitet von der Domäne β des Staphylococcenproteins A.
  - 8. Polypeptid nach Anspruch 5, worin die dritte Domäne eine Aminosäuresequenz umfasst, die an einen Zelloberflächenrezeptor bindet.
- 9. Polypeptid nach Anspruch 8, worin die dritte Domäne der insulinähnliche Wachstumsfaktor-1 (IFG-1 = insulin-like growth factor-1) ist.
  - 10. Polypeptid nach einem der vorstehenden Ansprüche, umfassend eine leichte Kette eines Botulinustoxins oder ein Fragment oder eine Variante einer leichten Kette eines Botulinustoxins und einen als H<sub>N</sub>-bezeichneten Teil der schweren Kette eines Botulinustoxins.
  - 11. Polypeptid nach Anspruch 10, worin eines oder beide von (a) der leichten Kette des Toxins oder des Fragments oder der Variante der leichten Kette des Toxins und (b) des Teils der schweren Kette des Toxins vom Botulinustoxin Typ A sind.
- 40 12. Polypeptid nach Anspruch 11, worin die Variante der leichten Kette des Botulinustoxins Typ A an Rest 2 ein Glutamat, an Rest 26 ein Lysin und an Rest 27 ein Tyrosin aufweist.
- 13. Polypeptid nach Anspruch 10, worin eines oder beide von (a) der leichten Kette des Toxins oder des Fragments oder der Variante der leichten Kette des Toxins und (b) des Teils der schweren Kette des Toxins vom Botulinustoxin
  45
  - 14. Polypeptid nach einem der Ansprüche 1-9, umfassend eine leichte Kette eines Botulinustoxins oder ein Fragment oder eine Variante einer leichten Kette eines Botulinustoxins und mindestens 100 N-terminale Aminosäuren einer schweren Kette eines Botulinustoxins.
  - 15. Polypeptid nach Anspruch 14, umfassend eine leichte Kette eines Botulinustoxins Typ B oder ein Fragment oder eine Variante davon und 107 N-terminale Aminosäuren einer schweren Kette eines Botulinustoxins Typ B.
  - Polypeptid nach Anspruch 11 oder 12, umfassend mindestens 423 der N-terminalen Aminosäuren der schweren Kette des Botulinustoxins Typ A.
    - Polypeptid gemäß Anspruch 16, umfassend eine leichte Kette eines Botulinustoxins Typ A und 423 N-terminale Aminosäuren der schweren Kette eines Botulinustoxins Typ A.

- 18. Polypeptid gemäß Anspruch 16, umfassend eine Variante der leichten Kette des Botulinustoxins Typ A, worin Rest 2 ein Glutamat, Rest 26 ein Lysin und Rest 27 ein Tyrosin ist, und 423 N-terminale Aminosäuren einer schweren Kette eines Botulinustoxins Typ A.
- Polypeptid nach Anspruch 13, umfassend mindestens 417 der N-terminalen Aminosäuren der schweren Kette des Botulinustoxins Typ B.

10

20

35

40

- Polypeptid nach Anspruch 19, umfassend eine leichte Kette eines Botulinustoxins Typ B und 417 N-terminale Aminosäuren einer schweren Kette eines Botulinustoxins Typ B.
- Polypeptid nach einem der Ansprüche 10-20, dem ein als H<sub>C</sub>-bezeichneter Teil einer schweren Kette eines Botulinustoxins fehlt
- 22. Polypeptid nach einem der Ansprüche 1-4, worin die zweite Domäne nicht an Zelloberflächenrezeptoren binden kann.
  - 23. Polypeptid nach einem der vorstehenden Ansprüche, umfassend eine Variante eines Clostridientoxins und zusätzlich umfassend eine Stelle zur Spaltung durch ein proteolytisches Enzym, welche Spaltstelle im nativen Toxin nicht vorhanden ist.
  - 24. Polypeptid nach Anspruch 23, umfassend eine Variante der leichten Kette eines Clostridientoxins und zusätzlich umfassend eine Stelle zur Spaltung durch ein proteolytisches Enzym, worin die Spaltstelle in der leichten Kette des nativen Toxins nicht vorhanden ist.
- 25. Polypeptid nach Anspruch 23 oder 24, umfassend eine Variante des H<sub>N</sub>-Teils einer schweren Kette eines Clostridientoxins und zusätzlich umfassend eine Stelle zur Spaltung durch ein proteolytisches Enzym, welche Spaltstelle in dem nativen H<sub>N</sub>-Teil der schweren Kette des Toxins nicht vorhanden ist.
- 26. Polypeptid nach Anspruch 23, 24 oder 25, erhältlich durch Modifikation einer das Polypeptid kodierenden DNS, um ein oder mehrere Nukleotide einzuführen, welche für die Spaltstelle kodieren.
  - 27. Fusionsprotein, umfassend eine Fusion von (a) einem Polypeptid nach einem der Ansprüche 1-26 mit (b) einem zweiten Polypeptid, bei dem es sich um ein Polypeptid oder Oligopeptid handelt, das an eine Affinitätsmatrix bindet, um die Reinigung des Fusionsproteins unter Verwendung der Matrix zu ermöglichen.
  - 28. Fusionsprotein gemäß Anspruch 27, worin das zweite Polypeptid an eine Chromatographiesäule bindet.
  - 29. Fusionsprotein nach Anspruch 28, worin die zweite Chromatographiesäule eine Affinitätsmatrix von Glutathionsepharose ist.
  - 30. Fusionsprotein nach Anspruch 27, 28 oder 29, worin eine spezifische Proteasespaltstelle zwischen dem ersten und zweiten Polypeptid eingebaut ist, worin die Proteasestelle die proteolytische Trennung des ersten und zweiten Polypeptids ermöglicht.
- 31. Zusammensetzung, umfassend ein Polypeptid nach einem der vorstehenden Ansprüche, worin die Zusammensetzung in vivo nicht toxisch ist.
  - 32. Zusammensetzung nach Anspruch 31 oder ein Polypeptid nach einem der Ansprüche 1-26 oder ein Fusionsprotein nach Anspruch 27, 28, 29 oder 30 zur Verwendung als Positivkontrolle in einem Toxintest.
  - 33. Zusammensetzung nach Anspruch 31 oder ein Polypeptid nach einem der Ansprüche 1-26 oder ein Fusionsprotein nach Anspruch 27, 28, 29 oder 30 zur Verwendung als ein Vaccin gegen ein Clostridientoxin.
- 34. Zusammensetzung gemäß Anspruch 31 oder ein Polypeptid nach einem der Ansprüche 1-26 oder ein Fusionsprotein nach Anspruch 27, 28, 29 oder 30 für die *in vivo*-Verwendung.
  - 35. Pharmazeutische Zusammensetzung, umfassend eine Zusammensetzung nach Anspruch 31, ein Polypeptid nach einem der Ansprüche 1-26 oder ein Fusionsprotein nach Anspruch 27, 28, 29 oder 30 in Kombination mit einem

pharmazeutisch annehmbaren Träger.

5

10

15

- 36. Nukleinsäure, kodierend ein Polypeptid oder ein Fusionsprotein nach einem der Ansprüche 1-30.
- Nukleinsäure nach Anspruch 36, kodierend ein Polypeptid oder ein Fusionsprotein und umfassend Nukleotide, kodierend die Reste 1-448 einer leichten Kette eines Botulinustoxins Typ A.
  - Nukleinsäure gemäß Anspruch 36 oder 37, umfassend Nukleotide, kodierend die Reste 1-423 einer H<sub>N</sub>-Domäne der schweren Kette eines Botulinustoxins Typ A.
  - Nukleinsäure nach Anspruch 36, kodierend ein Polypeptid oder ein Fusionsprotein und umfassend Nukleotide, kodierend die Reste 1-470 der leichten Kette eines Botulinustoxins Typ B.
- 40. Nukleinsäure nach Anspruch 36 oder 39, kodierend ein Polypeptid oder ein Fusionsprotein und umfassend Nukleotide, kodierend die Reste 1-417 einer H<sub>N</sub>-Domäne der schweren Kette des Botulinustoxins Typ B.
  - 41. Nukleinsäure nach einem der Ansprüche 36-40, umfassend Nukleotide, kodierend eine Restriktionsendonucleasespaltstelle, die in der nativen Sequenz des Clostridientoxins nicht vorhanden ist.
- 42. Nukleinsäure nach Anspruch 41, erhältlich durch Modifikation eines Nukleotids, kodierend ein Polypeptid oder Fusionsprotein nach einem der Ansprüche 1-30, um die Spaltstelle einzuführen.
  - 43. DNS nach einem der Ansprüche 36-42.
- 25 **44.** DNS, ausgewählt unter den SEQ IDs der Nummern 1, 3, 5, 9, 11, 13, 15, 17, 19, 21, 27 und 28.
  - 45. Verfahren zur Herstellung eines Polypeptids nach einem der Ansprüche 1-26, umfassend das Exprimieren einer Nukleinsäure nach einem der Ansprüche 36-44 in einer Wirtszelle und die Gewinnung des Polypeptids.
- 46. Verfahren der Herstellung eines Polypeptids nach einem der Ansprüche 1-26, umfassend das Exprimieren einer Nukleinsäure, kodierend ein Fusionsprotein nach Anspruch 27, 28, 29 oder 30 in einer Wirtszelle, das Reinigen des Fusionsproteins mittels Elution des Fusionsproteins durch eine Affinitätsmatrix, angepasst an das Zurückhalten des Fusionsproteins, und Eluieren eines Liganden durch die Matrix, der an das Verdrängen des Fusionsproteins angepasst ist, und das Gewinnen des Fusionsproteins.
  - 47. Verfahren zur Herstellung nach Anspruch 45 oder 46, worin die Nukleinsäure DNS ist.
  - 48. Zelle, exprimierend ein Polypeptid oder Fusionsprotein nach einem der Ansprüche 1-30.

### Revendications

40

45

50

- 1. Polypeptide comprenant un premier et un second domaines, dans lequel
  - ledit premier domaine est une chaîne légère de neutoroxine clostridienne ou un fragment ou un variant de ladite chaîne légère de neutoroxine clostridienne, ledit premier domaine étant capable de cliver une ou plusieurs protéines associées à une membrane vésiculaire ou plasmique et essentielles pour l'exocytose,
  - ledit deuxième domaine est une portion  $H_N$  de chaîne lourde de neurotoxine clostridienne ou un fragment ou un variant de la dite portion  $H_N$  de chaîne lourde de neutoroxine clostridienne, ledit deuxième domaine étant capable de (i) transloquer le polypeptide dans une cellule ou (ii) accroître la solubilité du polypeptide en comparaison avec la solubilité du premier domaine à lui seul ou (iii) à la fois transloquer le polypeptide dans une cellule et accroître la solubilité du polypeptide en comparaison avec la solubilité du premier domaine à lui seul,
  - une portion intacte désignée  $H_C$  d'une chaîne lourde de neurotoxine clostridienne étant absente dudit deuxième domaine, et
    - ledit polypeptide est un polypeptide à chaîne unique.
- 2. Polypeptide selon la revendication 1, dans lequel la toxine clostridienne est une toxine botulinique.
- 3. Polypeptide selon la revendication 1 ou la revendication 2, dans lequel le premier domaine possède une activité

endopeptidase spécifique d'un substrat choisi parmi un ou plusieurs de SNAP-25, synaptobrévine/VAMP et syntaxine.

- Polypeptide selon la revendication 1, dans lequel la portion H<sub>N</sub> de chaîne lourde de toxine clostridienne provient d'une toxine botulinique.
  - 5. Polypeptide selon l'une quelconque des revendications 1 à 4 comprenant en outre un troisième domaine liant le polypeptide à une cellule, en liant le troisième domaine directement à une cellule ou en liant le troisième domaine à un ligand ou à des ligands qui se lient à une cellule.
  - Polypeptide selon la revendication 5, dans lequel ledit troisième domaine est destiné à lier le polypeptide à une immunoglobuline.
- Polypeptide selon la revendication 6, dans lequel ledit troisième domaine est un domaine synthétique de liaison à l'IgG à séquences répétées en tandem dérivé du domaine bêta de la protéine A staphylococcique.

10

25

30

35

40

45

- 8. Polypeptide selon la revendication 5, dans lequel ledit troisième domaine comprend une séquence d'acides aminés qui se lie à un récepteur de surface cellulaire.
- Polypeptide selon la revendication 8, dans lequel ledit troisième domaine est le facteur de croissance insulinomimétique 1 (IGF-1).
  - 10. Polypeptide selon l'une quelconque des revendications précèdentes, comprenant une chaîne légère de toxine botulinique ou un fragment ou un variant d'une chaîne légère de toxine botulinique et une portion désignée H<sub>N</sub> d'une chaîne lourde de toxine botulinique.
    - 11. Polypeptide selon la revendication 10, dans lequel l'un ou les deux parmi (a) la chaîne légère de toxine ou le fragment ou le variant de chaîne légère de toxine et (b) la portion de la chaîne lourde de toxine, sont ceux d'une toxine botulinique de type A.
    - 12. Polypeptide selon la revendication 11, dans lequel le variant de chaîne légère de toxine botulinique de type A possède un glutamate au résidu 2, une lysine au résidu 26, et une tyrosine au résidu 27.
  - 13. Polypeptide selon la revendication 10, dans lequel l'un ou les deux parmi (a) la chaîne légère de toxine ou le fragment ou le variant de chaîne légère de toxine et (b) la portion de la chaîne lourde de toxine, sont ceux d'une toxine botulinique de type B.
  - 14. Polypeptide selon l'une quelconque des revendications 1 à 9 comprenant une chaîne légère de toxine botulinique ou un fragment ou une variante d'une chaîne légère de toxine botulinique et au moins 100 acides aminés Nterminaux d'une chaîne lourde de toxine botulinique.
  - 15. Polypeptide selon la revendication 14 comprenant une chaîne légère de toxine botulinique de type B, ou un fragment ou un variant de celle-ci, et 107 acides aminés d'extrémité N-terminale d'une chaîne lourde de toxine botulinique de type B.
  - 16. Polypeptide selon la revendication 11 ou la revendication 12 comprenant au moins 423 des acides aminés N-terminaux d'une chaîne lourde de toxine botulinique de type A.
- 17. Polypeptide selon la revendication 16 comprenant une chaîne légère de toxine botulinique de type A et 423 acides aminés N-terminaux d'une chaîne lourde de toxine botulinique de type A.
  - 18. Polypeptide selon la revendication 16 comprenant un variant de chaîne légère de toxine botulinique de type A dans lequel le résidu 2 est un glutamate, le résidu 26 est une lysine et le résidu 27 est une tyrosine, et 423 acides aminés N-terminaux d'une chaîne lourde de toxine botulinique de type A.
  - Polypeptide selon la revendication 13 comprenant au moins 417 des acides aminés N-terminaux de la chaîne lourde de toxine botulinique de type B.

- 20. Polypeptide selon la revendication 19 comprenant au moins une chaîne légère de toxine botulinique de type B et 417 acides aminés N-terminaux d'une chaîne lourde de toxine botulinique de type B.
- 21. Polypeptide selon l'une quelconque des revendication 10 à 20 dans lequel il manque une portion désignée H<sub>C</sub> d'une chaîne lourde de toxine botulinique est absente.

5

20

45

- 22. Polypeptide selon l'une quelconque des revendications 1 à 4, dans lequel le deuxième domaine n'est pas capable de se lier à des récepteurs de surface cellulaires.
- 23. Polypeptide selon l'une quelconque des revendications précèdentes comprenant un variant d'une toxine clostridienne et comprenant en outre un site pour le clivage par une enzyme protéolytique, lequel site de clivage n'est pas présent dans la toxine native.
- 24. Polypeptide selon la revendication 23 comprenant un variant d'une chaîne légère de toxine clostridienne et comprenant en outre un site pour le clivage par une enzyme protéolytique, lequel site de clivage n'est pas présent dans la chaîne légère de la toxine native.
  - 25. Polypeptide selon la revendication 23 ou la revendication 24 comprenant un variant d'une portion H<sub>N</sub> de chaîne lourde de toxine clostridienne et comprenant en outre un site pour le clivage par une enzyme protéolytique, lequel site de clivage n'est pas présent dans la portion H<sub>N</sub> de la chaîne lourde de la toxine native.
  - 26. Polypeptide selon la revendication 23, 24 ou 25 pouvant être obtenu par modification d'un ADN codant pour le polypeptide de manière à y introduire un ou plusieurs nucléotides codant pour le site de clivage.
- 27. Protéine de fusion formée par la fusion de (a) un polypeptide selon l'une quelconque des revendications 1 à 26 avec (b) un deuxième polypeptide qui est un polypeptide ou un oligopeptide qui se lie à une matrice d'affinité de manière à permettre une purification de la protéine de fusion en utilisant ladite matrice.
- 28. Protéine de fusion selon la revendication 27, dans laquelle ledit deuxième polypeptide se lie à une colonne de chromatographie.
  - 29. Protéine de fusion selon la revendication 28, dans laquelle ladite colonne de chromatographie est une matrice d'affinité de sépharose glutathione.
- 30. Protéine de fusion selon la revendication 27, 28 ou 29, dans laquelle un site spécifique de clivage par protéase est incorporé entre les premier et deuxième polypeptides, ledit site de clivage par protéase permettant la séparation protéolytique des premier et deuxième polypeptides.
- 31. Composition comprenant une polypeptide selon l'un quelconque des revendications précèdentes, ladite composition étant non toxique in vivo.
  - 32. Composition selon la revendication 31 ou polypeptide selon l'une quelconque des revendications 1 à 26 ou protéine de fusion selon la revendication 27, 28, 29 ou 30 destinés à une utilisation comme témoin positif dans un dosage de toxine.
  - 33. Composition selon la revendication 31 ou polypeptide selon l'une quelconque des revendications 1 à 26 ou protéine de fusion selon la revendication 27, 28, 29 ou 30 destinés à une utilisation comme vaccin contre une toxine clostridienne.
- 34. Composition selon la revendication 31 ou polypeptide selon l'une quelconque des revendications 1 à 26 ou protéine de fusion selon la revendication 27, 28, 29 ou 30 destinés à une utilisation in vivo.
  - 35. Composition pharmaceutique comprenant une composition selon la revendication 31, un polypeptide selon l'une quelconque des revendications 1 à 26, ou une protéine de fusion selon la revendication 27, 28, 29 ou 30, en combinaison avec un support physiologiquement acceptable.
  - 36. Acide nucléique codant pour un polypeptide ou une protéine de fusion selon l'une quelconque des revendications 1 à 30.

- 37. Acide nucléique codant pour un polypeptide ou une protéine de fusion selon la revendication 36 et comprenant des nucléotides codant pour des résidus 1 à 448 d'une chaîne légère de toxine botulinique de type A.
- 38. Acide nucléique selon la revendication 36 ou la revendication 37 comprenant des nucléotides codant pour des résidus 1 à 423 d'un domaine H<sub>N</sub> d'une chaîne lourde de toxine botulinique de type A.
  - 39. Acide nucléique codant pour un polypeptide ou une protéine de fusion selon la revendication 36 et comprenant des nucléotides codant pour des résidus 1 à 470 d'une chaîne légère de toxine botulinique de type B.
- 40. Acide nucléique codant pour un polypeptide ou une protéine de fusion selon la revendication 36 ou la revendication 39 comprenant des nucléotides codant pour des résidus 1 à 417 d'un domaine H<sub>N</sub> de chaîne lourde de toxine botulinique de type B.
- 41. Acide nucléique selon l'une quelconque des revendications 36 à 40 comprenant des nucléotides codant pour une site de clivage par endonucléase de restriction non présent dans la séquence de toxine clostridienne native.
  - **42.** Acide nucléique selon la revendication 41 susceptible d'être obtenu par modification d'un nucléotide codant pour un polypeptide ou une protéine de fusion selon l'une quelconque des revendications 1 à 30 de sorte à introduire ledit site de clivage.
  - 43. ADN selon l'une quelconque des revendications 36 à 42.

5

20

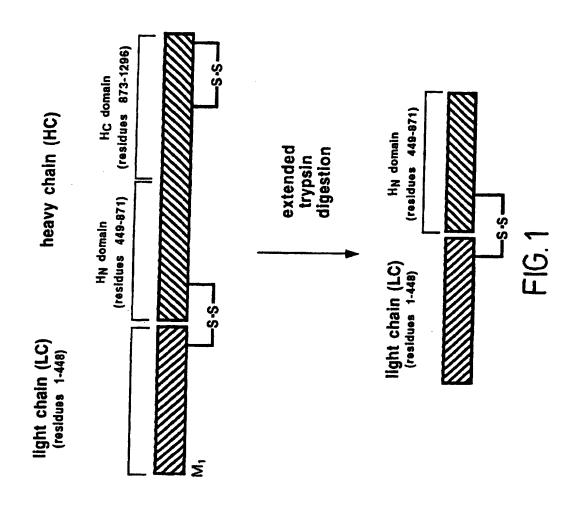
40

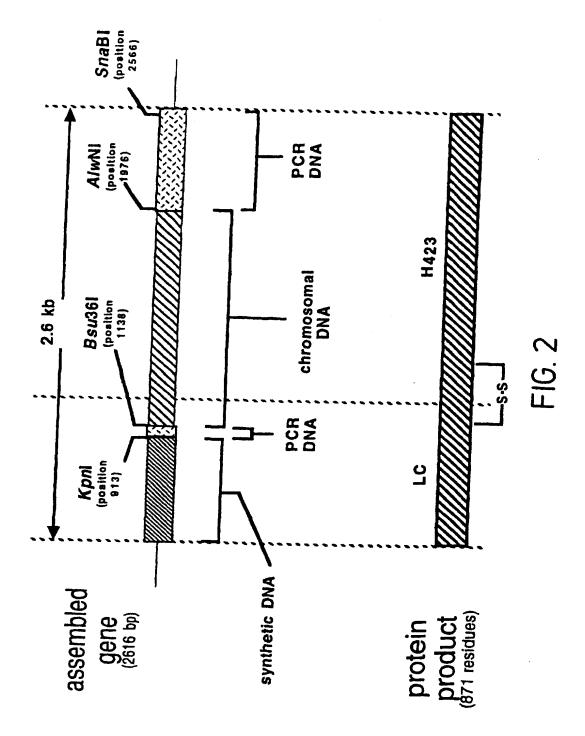
45

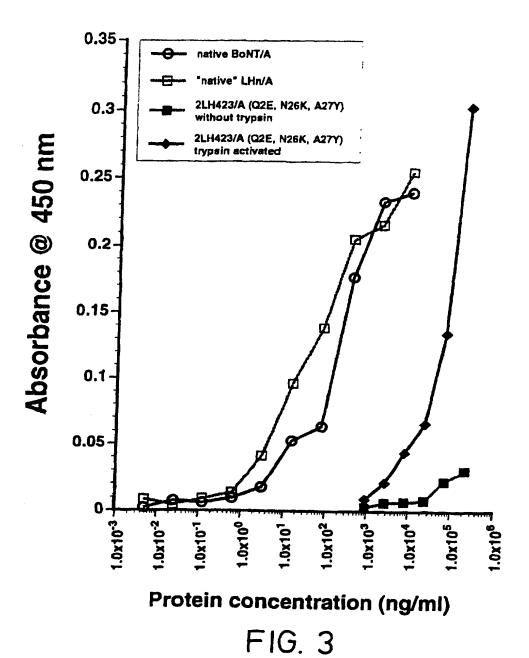
50

55

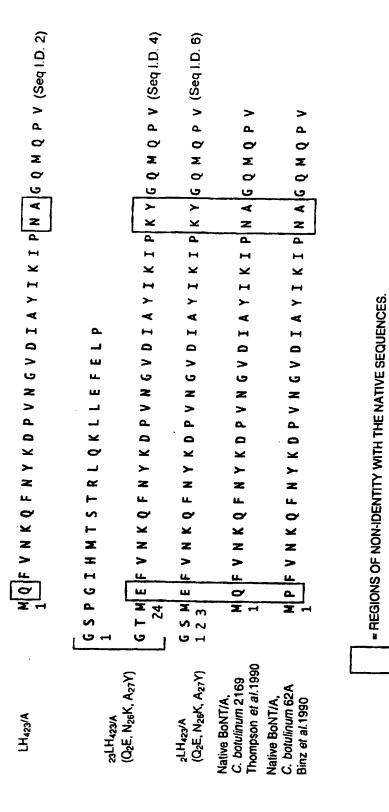
- **44**. ADN choisi parmi SEQ ID Nos 1, 3, 5, 9, 11, 13, 15, 17, 19, 21, 27 et 28.
- 45. Procédé de fabrication d'un polypeptide selon l'une quelconque des revendications 1 à 26 comprenant l'expression dans une cellule hôte d'un acide nucléique selon l'une quelconque des revendications 36 à 44 et la récupération polypeptide.
- 46. Procédé de fabrication d'un polypeptide selon l'une quelconque des revendications 1 à 26 comprenant l'expression dans une cellule hôte d'un acide nucléique codant pour une protéine de fusion selon la revendication 27, 28, 29 ou 30, la purification de la protéine de fusion par élution de la protéine de fusion à travers une matrice d'affinité adaptée pour retenir la protéine de fusion et par élution à travers ladite matrice d'un ligand adapté pour déplacer la protéine de fusion, et la récupération la protéine de fusion.
- 47. Procédé de fabrication selon les revendications 45 ou 46, dans lequel l'acide nucléique est de l'ADN.
  - 48. Cellule exprimant un polypeptide ou une protéine de fusion selon l'une quelconque des revendications 1 à 30.



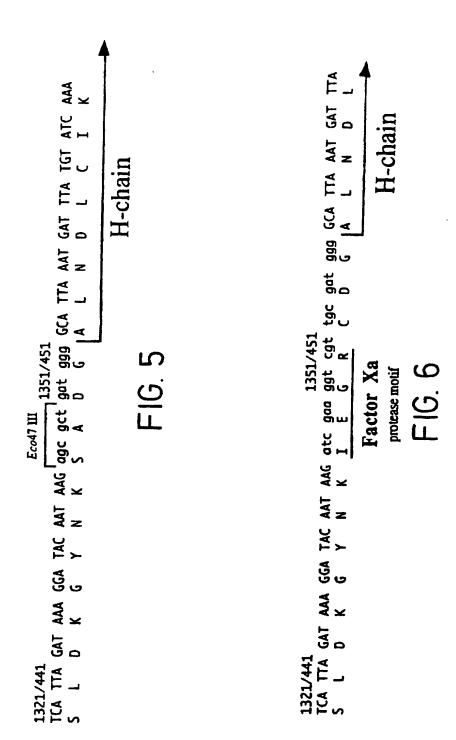




134



. C



GAC AGG D R **∀** 0 0 CGG AGC R S 666 TAT 6 Y GAG CTC AAG CCT GCC AAG L K P A K GCT A 999 AAG K ر ا ا GTG GAT V D 2587/863
TAC GTA GAT A
Y V D N
2647/883
CCG GAG ACG C
P E T L
2707/903
GCC TTT TAT 1
G F Y F

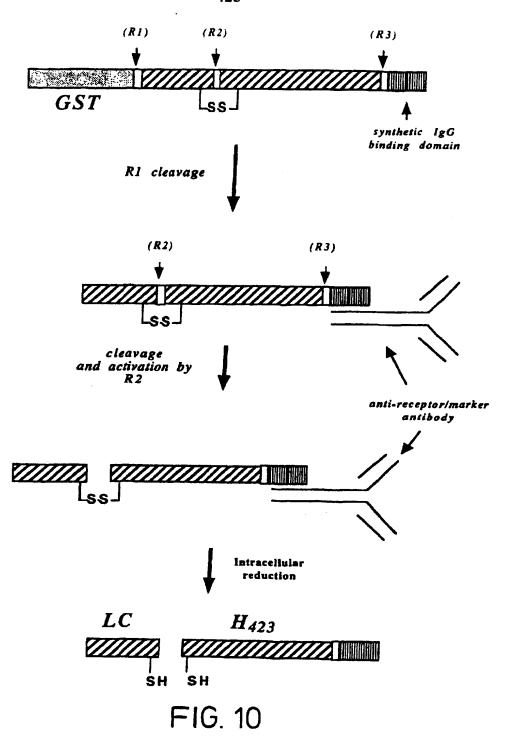
F16. 7

ACA CAT AAT T H N AAG K GAT 2617/873
A TCT ACA TTT ACT GAA TAT A
S T F T E Y I
2677/893
T TCA GGC TAT CAA TCT GAT A
S G Y Q S D I E J ¥, CAA ATA Q I AGA R AGA R AAA K 2587/863 TAC GTA GAT / V D 2647/883 TCT AAA GTT / S K V I

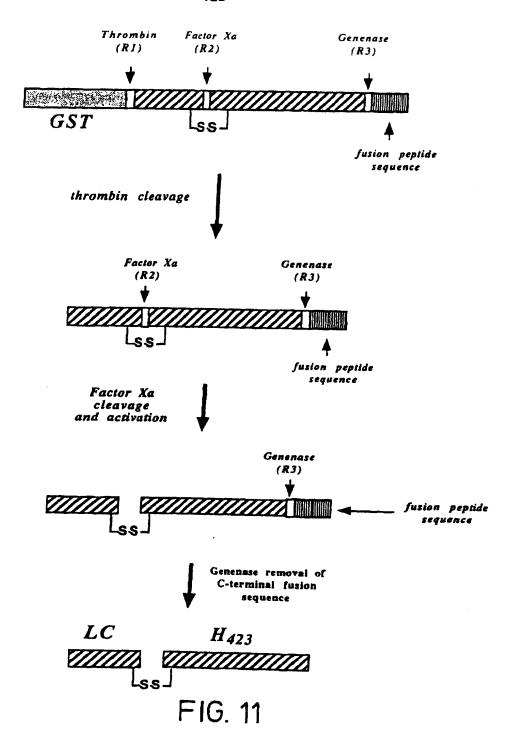
F16.8

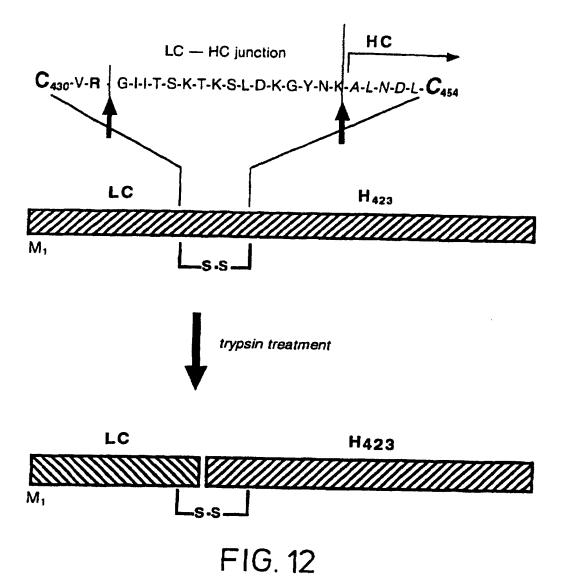
N N N K K CGA CGA Q Q GCC GAA E GAA E AAC N AAC GCT A 2617/873
TTT ACT GAA
F T E
2677/893
GAT GAA GCC
D E A
2737/913
CAT TTA CCT
H L P
2797/933
CCA AGC CAA
P S Q
2857/953
AAA GTA GAC
K V D
2917/973
CCT AAC TTA
P N L
2977/993
CAA AGC GCT
Q S A
3037/1013 A H GAC S S CAA Q ATC I ¥ E GAG E AAA K CAG TAT Y TT A GAT AGT S CTA AAT L N **₹**~ 900 A 

# LH<sub>423</sub>/A



# LH<sub>423</sub>/A





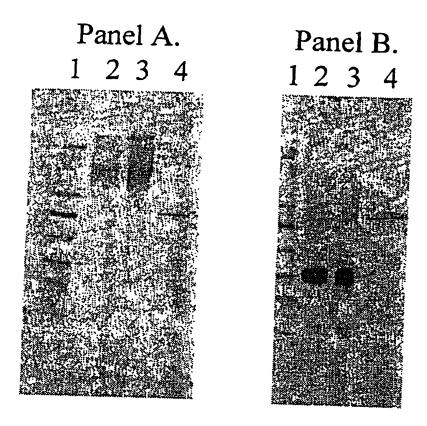


FIG. 13